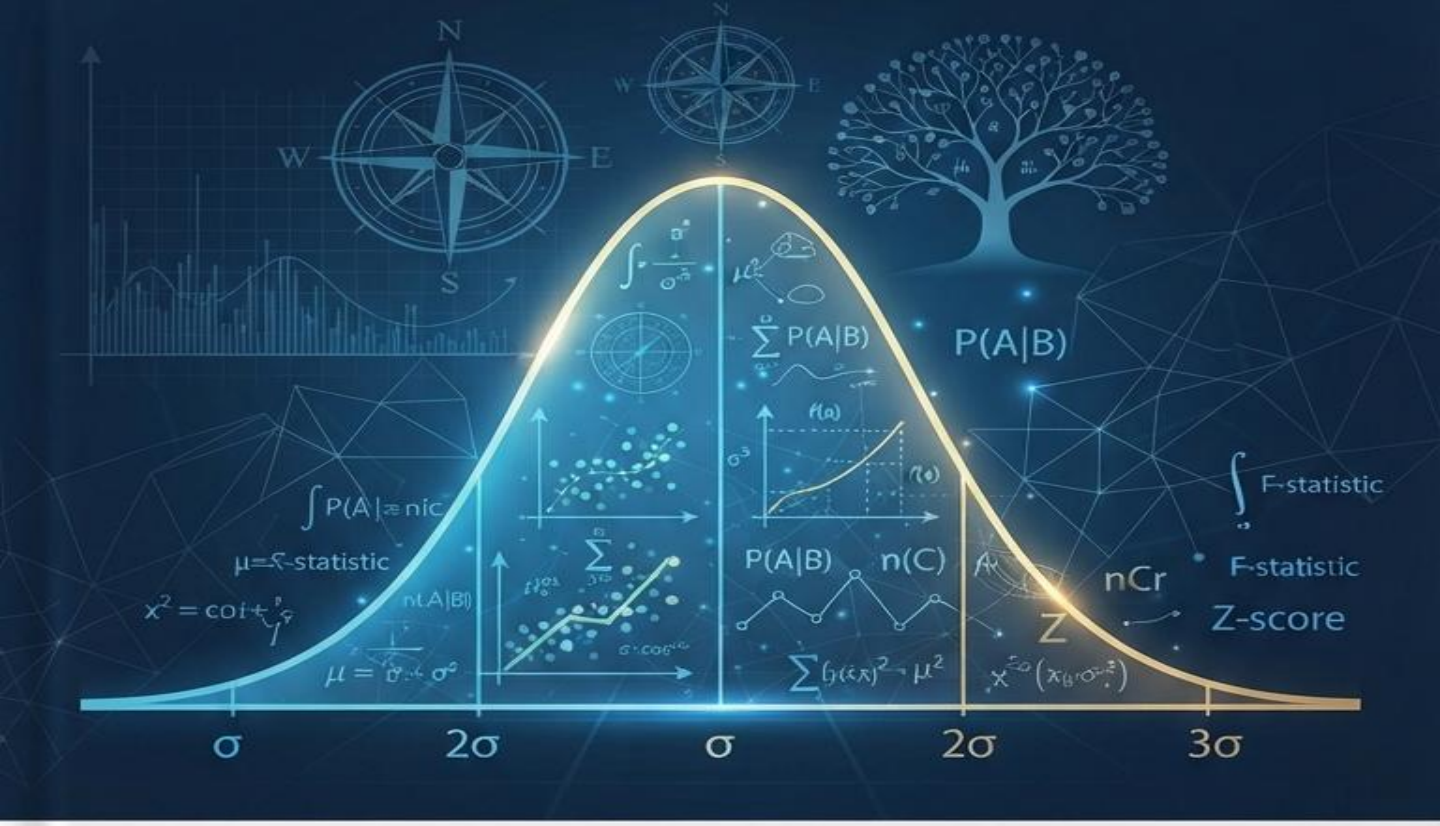
Probability and Statistical Inference

ဖြစ်တန်စွမ်းနှင့် စာရင်းအင်းဆိုင်ရာ ကောက်ချက်ချခြင်း



ဘာသာပြန် - နေအောင်

A Comprehensive Academic Guide for Students and Professionals

Academic Press
Myanmar

Probability and Statistical Inference (9th Edition)

အခန်း (၁) ဖြစ်တန်စွမ်းသီအိုရီ (Probability) ၏ အခြေခံများ

ဒီအခန်းမှာ ဖြစ်တန်စွမ်းရဲ့ ဂုဏ်သတ္တိတွေ၊ ရေတွက်နည်းစနစ်တွေ၊ အခြေအနေပေါ်မူတည်တဲ့ ဖြစ်တန်စွမ်း (Conditional Probability)၊ အမှီအခိုကင်းသော ဖြစ်ရပ်များ (Independent Events) နဲ့ Bayes' Theorem တို့ကို ရှင်းပြထားပါတယ်။

အခန်း (၂) အပိုင်းအခြားသတ်မှတ်နိုင်သော ဖြန့်ဖြူးမှုများ (Discrete Distributions)

ဒီအခန်းမှာ အပိုင်းအခြားသတ်မှတ်နိုင်တဲ့ ကျပန်းကိန်းရှင်များ (Random Variables)၊ သခ်ၤာမျှော်မှန်းချက် (Mathematical Expectation)၊ Binomial၊ Negative Binomial နဲ့ Poisson ဖြန့်ဖြူးမှုတွေအကြောင်း အသေးစိတ် ရှင်းပြထားပါတယ်။

အခန်း (၃) အဆက်မပြတ် ဖြန့်ဖြူးမှုများ (Continuous Distributions)

ဒီအခန်းမှာ အဆက်မပြတ် ကျပန်းကိန်းရှင်များ၊ Exponential၊ Gamma၊ Chi-Square နဲ့ Normal ဖြန့်ဖြူးမှုတွေအကြောင်း စတင် မိတ်ဆက်ထားပါတယ်။

အခန်း (၄) နှစ်ခုမြောက် ဖြန့်ဖြူးမှုများ (Bivariate Distributions)

ဒီအခန်းမှာ ကိန်းရှင်နှစ်ခုပါတဲ့ ဖြန့်ဖြူးမှုတွေ၊ ဆက်စပ်မှုကိန်း (Correlation Coefficient)၊ အခြေအနေအရ ဖြန့်ဖြူးမှုတွေ (Conditional Distributions) နဲ့ Bivariate Normal Distribution အကြောင်း ရှင်းပြထားပါတယ်။

အခန်း (၅) ကျပန်းကိန်းရှင်များ၏ လုပ်ဆောင်ချက်များ၏ ဖြန့်ဖြူးမှု

ဒီအခန်းမှာ ကျပန်းကိန်းရှင်တစ်ခုနဲ့ နှစ်ခုရဲ့ အသွင်ပြောင်းလုပ်ဆောင်ချက်တွေ၊ Moment-Generating Function နည်းပညာ၊ Central Limit Theorem နဲ့ Chebyshev's Inequality တို့အကြောင်း ပါဝင်ပါတယ်။

အခန်း (၆) အမှတ်ခန့်မှန်းခြင်း (Point Estimation)

ဒီအခန်းမှာ ဖော်ပြချက်စာရင်းဇယား (Descriptive Statistics)၊ အစီအစဉ်စာရင်းဇယား (Order Statistics)၊ Maximum Likelihood Estimation၊ Sufficient Statistics နဲ့ Bayesian Estimation အကြောင်း စတင် ရှင်းပြထားပါတယ်။

အခန်း (၇) ကြားကာလခန့်မှန်းခြင်း (Interval Estimation)

ဒီအခန်းမှာ ပျမ်းမျှအတွက် ယုံကြည်မှုကြားကာလ (Confidence Intervals for Means)၊ အချိုးအစားအတွက် ယုံကြည်မှုကြားကာလ၊ နမူနာအရွယ်အစား သတ်မှတ်ခြင်းနဲ့ ပြန်လည်နမူနာယူခြင်းနည်းလမ်းများ (Resampling Methods) အကြောင်း ပါဝင်ပါတယ်။

အခန်း (၈) စာရင်းအင်းဆိုင်ရာ ယူဆချက်များကို စစ်ဆေးခြင်း

ဒီအခန်းမှာ ပျမ်းမျှတစ်ခုအကြောင်း စစ်ဆေးခြင်း၊ ပျမ်းမျှနှစ်ခု ညီမျှမှုကို စစ်ဆေးခြင်း၊ အချိုးအစားများအကြောင်း စစ်ဆေးခြင်း၊ Wilcoxon စစ်ဆေးခြင်းနဲ့ စစ်ဆေးမှုတစ်ခုရဲ့ ပါဝါ (Power of a Test) အကြောင်း ရှင်းပြထားပါတယ်။

အခန်း (၉) နောက်ထပ် စစ်ဆေးမှုများ

ဒီနောက်ဆုံးအခန်းမှာ Chi-Square Goodness-of-Fit စစ်ဆေးခြင်း၊ Contingency Tables၊ One-Factor နဲ့ Two-Way Analysis of Variance (ANOVA)၊ စက်မှုလုပ်ငန်းသုံး စာရင်းအင်းအရည်အသွေးထိန်းချုပ်ခြင်း (Statistical Quality Control) အကြောင်း ပါဝင်ပါတယ်။

အခန်း (၁) ဖြစ်တန်စွမ်းသီအိုရီ (Probability) ၏ အခြေခံများ

ဒီအခန်းရဲ့ ရည်ရွယ်ချက်က **ဖြစ်တန်စွမ်း (Probability)** ဆိုတာ ဘာလဲ၊
ဘယ်လိုတွက်ရမလဲဆိုတာကို အခြေခံကနေ စနစ်တကျ သင်ယူဖို့ပါ။ ဒီမှာ အဓိက သင်ယူရမှာက (၁) ဖြစ်တန်စွမ်းရဲ့ ဂုဏ်သတ္တိတွေ၊ (၂) ရေတွက်နည်းစနစ်တွေ၊ (၃) အခြေအနေအရ ဖြစ်တန်စွမ်း၊ (၄) အမှီအခိုကင်းတဲ့ ဖြစ်ရပ်တွေ၊ နဲ့ (၅) Bayes' Theorem တို့ပဲ ဖြစ်ပါတယ်။

၁.၁ ဖြစ်တန်စွမ်းရဲ့ ဂုဏ်သတ္တိများ (Properties of Probability)

ဒီအပိုင်းမှာ သခ်ၤာဘာသာရပ်မှာ ဖြစ်တန်စွမ်း (Probability) ဆိုတဲ့ စကားလုံးကို ဘယ်လို သတ်မှတ်သုံးစွဲသလဲဆိုတာ လေ့လာရမှာပါ။

- **အခြေခံ အယူအဆ**
    - **Random Experiment (ကျပန်းစမ်းသပ်မှု):** ရလဒ်ကို ကြိုမသိနိုင်တဲ့ စမ်းသပ်မှုမျိုး (ဥပမာ - အကြွေစေ့ ပစ်ခြင်း၊ အန်စာတုံး လှိမ့်ခြင်း)။
    - **Sample Space (S) (ရလဒ်အားလုံးရဲ့ အစု):** စမ်းသပ်မှုတစ်ခုမှာ ဖြစ်နိုင်သမျှ ရလဒ်အားလုံးကို စုစည်းထားတဲ့ အစု။ (ဥပမာ - အကြွေစေ့ပစ်ရင် S = {ခေါင်း, ပြင်})
    - **Event (A) (ဖြစ်ရပ်):** Sample Space ထဲက ရလဒ်အချို့ကို စုစည်းထားတဲ့ အစုငယ်တစ်ခု။ (ဥပမာ - ခေါင်းကျတဲ့ ဖြစ်ရပ် A = {ခေါင်း})
- **ဖြစ်တန်စွမ်းရဲ့ အခြေခံစည်းမျဉ်းများ**
  ဖြစ်တန်စွမ်း P(A) ဆိုတဲ့ တန်ဖိုးဟာ အောက်ပါ စည်းမျဉ်းတွေကို လိုက်နာရပါမယ်။
    1. P(A) ဟာ အမြဲတမ်း 0 ထက် ကြီးသည် သို့မဟုတ် ညီသည်။ (P(A) ≥ 0)
    2. Sample Space တစ်ခုလုံးရဲ့ဖြစ်တန်စွမ်းက 1 ဖြစ်သည်။ (P(S) = 1)

3. ဖြစ်ရပ်နှစ်ခု ထပ်မနေဘူးဆိုရင် (A နဲ့ B မှာ တူညီတဲ့အချက်မရှိဘူးဆိုရင်) သူတို့ပေါင်းစည်းတဲ့ဖြစ်ရပ်ရဲ့ဖြစ်တန်စွမ်းက သူတို့တစ်ခုချင်းစီရဲ့ဖြစ်တန်စွမ်းတွေပေါင်းလောက်ဖြစ်တယ်။ (P(A∪B) = P(A) + P(B))

- **အရေးကြီးတဲ့ သီအိုရမ်များ**
  ဒီစည်းမျဉ်းတွေကနေ ဆင်းသက်လာတဲ့ အရေးကြီးတဲ့ ဖော်မြူလာတချို့ကတော့...
  - **ဖြည့်စွက်ဖြစ်ရပ် (Complement):** ဖြစ်ရပ် A မဖြစ်ဖို့ ဖြစ်တန်စွမ်းက 1 အနုတ် A ဖြစ်တဲ့ ဖြစ်တန်စွမ်းပါ။ P(A') = 1 - P(A)
  - **ပေါင်းစည်းခြင်း (Union):** ဖြစ်ရပ် A ဒါမှမဟုတ် B ဖြစ်ဖို့ ဖြစ်တန်စွမ်းက P(A∪B) = P(A) + P(B) - P(A∩B) ။ ဒီမှာ P(A∩B) က A နဲ့ B နှစ်ခုလုံး ဖြစ်တဲ့ ဖြစ်တန်စွမ်းပါ။ (ဒီအချက်က နှစ်ခုလုံးဖြစ်တာကို နှစ်ခါမထည့်မိအောင် နုတ်ပေးရတာပါ)
- **ဥပမာ - ကတ်ပြားဆွဲခြင်း**
  ကတ် ၅၂ ရွက်ကနေ ဘုရင် (King) တစ်ရွက် ဆွဲဖို့ ဖြစ်တန်စွမ်းက ဘုရင် ၄ ရွက် ရှိတဲ့အတွက် 4/52 = 1/13 ဖြစ်ပါတယ်။

၁.၂ ရေတွက်နည်းစနစ်များ (Methods of Enumeration)

ဖြစ်တန်စွမ်းတွက်တဲ့အခါ ဖြစ်နိုင်တဲ့ ရလဒ်အရေအတွက် (Total Outcomes) နဲ့ လိုချင်တဲ့ ရလဒ်အရေအတွက် (Favorable Outcomes) ကို သိဖို့လိုပါတယ်။ အဲဒါကို ရေတွက်ဖို့ အောက်ပါနည်းလမ်းတွေကို သုံးပါတယ်။

- **မြောက်ခြင်းစည်းမျဉ်း (Multiplication Principle):**
  အလုပ် (A) ကို m နည်းနဲ့ လုပ်နိုင်ပြီး၊ အဲဒီနောက် အလုပ် (B) ကို n နည်းနဲ့ လုပ်နိုင်ရင် အလုပ် A နဲ့ B ကို တွဲပြီးလုပ်တဲ့ နည်းလမ်း အရေအတွက်က m × n ဖြစ်ပါတယ်။
  - **ဥပမာ:** အင်္ကျီ ၃ ထည်နဲ့ ဘောင်းဘီ ၂ ထည် ရွေးချင်ရင် ဝတ်စုံပေါင်း 3 × 2 = 6 မျိုး ဖြစ်နိုင်ပါတယ်။

- **ပြောင်းလဲအစီအစဉ် (Permutation):**
  အရာဝတ္ထု n ခုကို အစီအစဉ်တကျ စီထားတဲ့ နည်းလမ်းကို n! (n factorial) လို့ ခေါ်ပါတယ်။ (ဥပမာ - A, B, C ဆိုတဲ့ စာလုံး ၃ လုံးကို ဘယ်လိုမျိုးစီစဉ်လို့ရလဲဆိုတဲ့ အစီအစဉ်ပေါင်း 3! = 6 မျိုး)။
  အကယ်၍ n ခုထဲက r ခုကိုပဲ အစီအစဉ်တကျ ရွေးချင်ရင် နည်းလမ်းပေါင်းက nPr = n! / (n-r)! ဖြစ်ပါတယ်။
  - **ဥပမာ:** လူ ၁၀ ယောက်ထဲက ဥက္ကဋ္ဌ၊ ဒုဥက္ကဋ္ဌ ရွေးချင်ရင်
    အစီအစဉ်ပါသွားတဲ့အတွက် 10P2 = 10 × 9 = 90 မျိုး ရွေးလို့ရပါတယ်။
- **ပေါင်းစပ်ခြင်း (Combination):**
  အရာဝတ္ထု n ခုထဲက r ခုကို အစီအစဉ်မပါဘဲ ရွေးချင်တဲ့ နည်းလမ်းကို nCr = n! / [r! (n-r)!] လို့ တွက်ပါတယ်။
  - **ဥပမာ:** ကတ် ၅၂ ရွက်ထဲက ကတ် ၅ ရွက် ကိုင်မယ့် လက်တစ်ခုရဲ့ဖြစ်နိုင်ခြေ အရေအတွက်က 52C5 ဖြစ်ပါတယ်။ ဒါက ကတ်တွေရဲ့အစီအစဉ် မပါဘဲ ရောနှောမှုကိုပဲ တွက်တာပါ။

၁.၃ အခြေအနေအရ ဖြစ်တန်စွမ်း (Conditional Probability)

ဒါက "ဒီအခြေအနေအောက်မှာ ဒီဖြစ်ရပ် ဖြစ်နိုင်ခြေ ဘယ်လောက်လဲ" ဆိုတာ တွက်တာပါ။

- **အဓိပ္ပါယ်:** ဖြစ်ရပ် B ဖြစ်ပြီးဆိုတဲ့ အခြေအနေအောက်မှာ ဖြစ်ရပ် A ဖြစ်ဖို့ ဖြစ်တန်စွမ်းကို P(A|B) လို့ ရေးပြီး P(A|B) = P(A∩B) / P(B) လို့ တွက်ပါတယ်။ (P(B) က 0 ထက် ကြီးရပါမယ်)
- **ဥပမာ:** ပန်းအိုးထဲမှာ ဘောလုံးနီ ၃ လုံး၊ အပြာ ၇ လုံးရှိတယ်။ ပထမဆုံး ဘောလုံးနီတစ်လုံး ထုတ်ပြီးသွားရင် (ပြန်မထည့်ဘဲ) ဒုတိယအကြိမ် ဘောလုံးပြာထွက်ဖို့ ဖြစ်တန်စွမ်းက P(ပြာ | နီ) = 7/9 ဖြစ်ပါတယ်။ ဘာကြောင့်လဲဆိုတော့ နီတစ်လုံး ထွက်သွားတဲ့အတွက် ကျန်တဲ့ ဘောလုံးပေါင်း ၉ လုံးထဲမှာ အပြာ ၇ လုံးပဲ ရှိတော့လို့ပါ။

၁.၄ အမှီအခိုကင်းသော ဖြစ်ရပ်များ (Independent Events)

ဖြစ်ရပ် A ဖြစ်တာက ဖြစ်ရပ် B ရဲ့ဖြစ်တန်စွမ်းကို ပြောင်းလဲမသွားဘူးဆိုရင် အဲဒီဖြစ်ရပ်နှစ်ခုကို အမှီအခိုကင်းတယ်လို့ ခေါ်ပါတယ်။

- **သတ်မှတ်ချက်:** A နဲ့ B က လွတ်လပ်တယ်ဆိုရင် $P(A \cap B) = P(A) \times P(B)$ ဖြစ်ရပါမယ်။
- **ဥပမာ:** အကြွေစေ့ကို နှစ်ခါပစ်တယ်။ ပထမအကြိမ် ခေါင်းကျတာနဲ့ ဒုတိယအကြိမ် ခေါင်းကျတာဟာ အမှီအခိုကင်းပါတယ်။ ဘာကြောင့်လဲဆိုတော့ ပထမအကြိမ်ရဲ့ရလဒ်က ဒုတိယအကြိမ်ရဲ့ရလဒ်ကို သက်ရောက်မှုမရှိလို့ပါ။

၁.၅ Bayes' Theorem (ဘေးွစ်သီအိုရမ်)

ဒါက အခြေအနေအရ ဖြစ်တန်စွမ်းရဲ့ဆန့်ကျင်ဘက်ပုံစံလို့ ဆိုနိုင်ပါတယ်။ ရလဒ်တစ်ခုကို သိပြီးသားဖြစ်တဲ့အခါ အဲဒီရလဒ်ဖြစ်ဖို့ အကြောင်းရင်းခံတွေရဲ့ဖြစ်တန်စွမ်းကို ပြန်တွက်တာပါ။

- **ပုံသေနည်း:**
  အကြောင်းရင်းခံ $B_1, B_2, \ldots, B_n$ တွေရှိပြီး ရလဒ် A ကိုတွေ့ရတယ်ဆိုရင် အကြောင်းရင်းခံ B_k ကြောင့် A ဖြစ်တာလားဆိုတဲ့ နောက်ဆုံး ဖြစ်တန်စွမ်း (Posterior Probability) ကို တွက်တဲ့ ပုံသေနည်းပါ။
  $P(B_k \mid A) = [P(B_k) \times P(A \mid B_k)] / [P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \ldots + P(B_n)P(A|B_n)]$
- **ဥပမာ - စက်ရုံသုံးရုံမှ ချိုယွင်းပစ္စည်းများ**
  စက်ရုံ (I) က ၃၅% ထုတ်လုပ်ပြီး ၂% ချိုယွင်းတယ်။ စက်ရုံ (II) က ၂၅% ထုတ်လုပ်ပြီး ၁% ချိုယွင်းတယ်။ စက်ရုံ (III) က ၄၀% ထုတ်လုပ်ပြီး ၃% ချိုယွင်းတယ်။ အခု ချိုယွင်းပစ္စည်းတစ်ခု တွေ့ရင် ဒီပစ္စည်းက စက်ရုံ (III) ကထွက်လာတာ ဖြစ်နိုင်ခြေ ဘယ်လောက်လဲ။
  အဖြေက Bayes' Theorem သုံးရင် $P(III|D) = (0.4 \times 0.03) / (0.35 \times 0.02 + 0.25 \times 0.01 + 0.4 \times 0.03) = 120/215$ လို့ ရပါတယ်။

အချုပ်အားဖြင့် ဒီအခန်းက ဖြစ်တန်စွမ်းဆိုတဲ့ သဘောတရားကို စတင်မိတ်ဆက်ပြီး၊ ရေတွက်နည်းစနစ်တွေ၊ အခြေအနေအရ ဖြစ်တန်စွမ်း၊ လွတ်လပ်မှု နဲ့ Bayes' Theorem ဆိုတဲ့ အခြေခံကျတဲ့ အကြောင်းအရာ (၄) ခုကို သင်ကြားပေးထားတာ ဖြစ်ပါတယ်။ ဒါတွေဟာ နောက်အခန်းတွေမှာ သင်ယူရမယ့် စာရင်းအင်းဆိုင်ရာ ခန့်မှန်းချက်တွေရဲ့ အခြေခံအုတ်မြစ် ဖြစ်ပါတယ်။

အခန်း (၂) အပိုင်းအခြားသတ်မှတ်နိုင်သော ဖြန့်ဖြူးမှုများ (Discrete Distributions)

ဒီအခန်းရဲ့အဓိက ရည်ရွယ်ချက်က **အပိုင်းအခြားသတ်မှတ်နိုင်သော ကျပန်းကိန်းရှင်များ (Discrete Random Variables)** နဲ့ သူတို့ရဲ့ဖြန့်ဖြူးမှု (Distribution) တွေကို လေ့လာဖို့ပါ။ ဒီမှာ သင်ယူရမှာက (၁) Discrete Random Variable ဆိုတာ ဘာလဲ၊ (၂) Mathematical Expectation (သခ်ၤာမျှော်မှန်းချက်)၊ (၃) Mean နဲ့ Variance လိုမျိုး အထူးမျှော်မှန်းချက်တွေ၊ (၄) Binomial Distribution၊ (၅) Negative Binomial Distribution နဲ့ (၆) Poisson Distribution တို့ပဲ ဖြစ်ပါတယ်။

၂.၁ အပိုင်းအခြားသတ်မှတ်နိုင်သော ကျပန်းကိန်းရှင် (Random Variables of the Discrete Type)

ဒီအပိုင်းမှာ ကျပန်းကိန်းရှင် (Random Variable) ဆိုတာ ဘာလဲ၊ အပိုင်းအခြားသတ်မှတ်နိုင်တဲ့ ကျပန်းကိန်းရှင်ဆိုတာ ဘာလဲဆိုတာ စပြီး လေ့လာရမှာပါ။

- **ကျပန်းကိန်းရှင် (Random Variable) ဆိုတာ**
  စမ်းသပ်မှုတစ်ခုရဲ့ရလဒ်တွေကို ကိန်းဂဏန်းတွေနဲ့ ဖော်ပြတဲ့ နည်းလမ်းတစ်ခုပါ။ ဥပမာ - အကြွေစေ့ပစ်ရင် ခေါင်းကျရင် ၁၊ ပြင်ကျရင် ၀ လို့ သတ်မှတ်တာမျိုးပါ။

- **Discrete Random Variable ဆိုတာ**
  တန်ဖိုးတွေကို ရေတွက်လို့ရတဲ့ ကျပန်းကိန်းရှင်မျိုးပါ။ (ဥပမာ - အန်စာတုံးပေါ်က အရေအတွက် {1,2,3,4,5,6}) ဒီတန်ဖိုးတွေကို စာရင်းသွင်းလို့ရပြီး အကန့်အသတ်ရှိတဲ့ တန်ဖိုးတွေ ဒါမှမဟုတ် ရေတွက်လို့ရတဲ့ အဆုံးမရှိတန်ဖိုးတွေ ဖြစ်ပါတယ်။

- **Probability Mass Function (pmf) - ဖြစ်တန်စွမ်းဩဇာဆိုချက်**
  ဒါက Discrete Random Variable တစ်ခုရဲ့တန်ဖိုးတစ်ခုစီအတွက် ဖြစ်တန်စွမ်းကို သတ်မှတ်ပေးတဲ့ ဖော်မြူလာပါ။

  - ဂုဏ်သတ္တိ (၂) ခုရှိတယ်။
    ၁။ $f(x) > 0$ (ဖြစ်တန်စွမ်းက ၀ ထက်ကြီးတယ်)
    ၂။ အစုထဲက ဖြစ်တန်စွမ်းအားလုံးပေါင်းက 1 ဖြစ်တယ်။ $\Sigma f(x) = 1$

- **Cumulative Distribution Function (cdf) - အစုအပြုံဖြန့်ဖြူးမှုဖွင့်ဆိုချက်**
  ဒါက ကျပန်းကိန်းရှင်ရဲ့တန်ဖိုးဟာ သတ်မှတ်ထားတဲ့ တန်ဖိုး x ထက် ငယ်တဲ့ ဒါမှမဟုတ် ညီတဲ့ ဖြစ်တန်စွမ်းကို ပြောတာပါ။ F(x) = P(X ≤ x)

- **ဥပမာ - အန်စာတုံးနှစ်ခုရဲ့အများဆုံးအရေအတွက်**
  အန်စာတုံး ၄ မျက်နှာပါတဲ့အန်စာတုံး နှစ်ခု လှိမ့်တယ်။ X က ရလဒ်နှစ်ခုထဲက အများဆုံးအရေအတွက်ဆိုရင် X ရဲ့ pmf က f(x) = (2x-1)/16 ဖြစ်ပြီး x = 1,2,3,4 အတွက် ဖြစ်တန်စွမ်းတွေက 1/16, 3/16, 5/16, 7/16 ဖြစ်ပါတယ်။

- **Hypergeometric Distribution - ဟိုက်ပါဂျီအိုမက်ထရစ်ဖြန့်ဖြူးမှု**
  ဒါက ပစ္စည်းအစုတစ်ခုထဲက ပစ္စည်းအချို့ကို ပြန်မထည့်ဘဲ ရွေးတဲ့အခါ သုံးတဲ့ ဖြန့်ဖြူးမှုပါ။
  ဥပမာ - ဘောလုံး ၁၀၀ ထဲမှာ အနီ ၂၀ ရှိတယ်။ အဲဒီထဲက ၅ လုံး ရွေးရင် အနီဘယ်နှစ်လုံးပါမလဲဆိုတဲ့ ဖြစ်တန်စွမ်းကို တွက်တာပါ။ ဒီဖြန့်ဖြူးမှုရဲ့ pmf က f(x) = [C(20,x) * C(80,5-x)] / C(100,5) ဖြစ်ပါတယ်။

၂.၂ သင်္ချာမျှော်မှန်းချက် (Mathematical Expectation)

ဒီအပိုင်းမှာ ကျပန်းကိန်းရှင်တစ်ခုရဲ့ "ပျမ်းမျှတန်ဖိုး" ကို ဘယ်လိုတွက်သလဲဆိုတာ လေ့လာရမှာပါ။

- **အဓိပ္ပါယ်**
  Discrete Random Variable X ရဲ့ pmf က f(x) ဆိုရင် X ရဲ့မျှော်မှန်းချက် (Expected Value) က E(X) = Σ [x * f(x)] ဖြစ်ပါတယ်။ ဒါက တန်ဖိုးတစ်ခုစီနဲ့ သူ့ရဲ့ဖြစ်တန်စွမ်းကို မြှောက်ပြီး အကုန်ပေါင်းတာပါ။
  ဒါကို ပျမ်းမျှ (Mean) လို့လည်း ခေါ်ပြီး μ နဲ့ ဖော်ပြလေ့ရှိပါတယ်။

- **ဥပမာ - ဂိမ်းတစ်ခုရဲ့ပျမ်းမျှဆုကြေး**
  အန်စာတုံးတစ်ခုကို လှိမ့်ပြီး
  - ၁,၂,၃ ကျရင် ၁ ဒေါ်လာ ရမယ်။
  - ၄,၅ ကျရင် ၂ ဒေါ်လာ ရမယ်။

- ၆ ကျရင် ၃ ဒေါ်လာ ရမယ်။
  ဆိုရင် မျှော်မှန်းထားတဲ့ ပျမ်းမျှဆုကြေးက E(X) = 1*(3/6) + 2*(2/6) + 3*(1/6) = 10/6 = 1.67 ဒေါ်လာ ဖြစ်ပါတယ်။

- **ဂုဏ်သတ္တိများ**
  - E(c) = c (ကိန်းသေတစ်ခုရဲ့ မျှော်မှန်းချက်က သူ့ဘာသာသူပါ)
  - E(cX) = cE(X) (ကိန်းသေနဲ့ မြှောက်ထားရင် အပြင်ကို ထုတ်လို့ရတယ်)
  - E(X+Y) = E(X) + E(Y) (ပေါင်းတာကို သပ်သပ်ခွဲလို့ရတယ်)

၂.၃ အထူးသဒ္ဓါမျှော်မှန်းချက်များ (Special Mathematical Expectations)

ဒီအပိုင်းမှာ Mean အပြင် တခြား အရေးကြီးတဲ့ ကိန်းဂဏန်းတွေဖြစ်တဲ့ Variance နဲ့ Moment-Generating Function (mgf) တို့ကို လေ့လာရမှာပါ။

- **Variance (ကွဲလွဲမှု)**
  ဒါက ကျပန်းကိန်းရှင်ရဲ့တန်ဖိုးတွေက ပျမ်းမျှကနေ ဘယ်လောက်ကွာဝေးသလဲဆိုတဲ့ အတိုင်းအတာပါ။
  - $Var(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2$
  - Standard Deviation (စံသွေဖည်မှု) က Variance ရဲ့နှစ်ထပ်ကိန်းရင်းပါ။ $\sigma = \sqrt{Var(X)}$
  - **ဥပမာ - အန်စာတုံးတစ်ခု** ရဲ့ Mean က ၃.၅ ဖြစ်ပြီး Variance က $(1^2+2^2+3^2+4^2+5^2+6^2)/6 - (3.5)^2 = 35/12$ ဖြစ်ပါတယ်။

- **Moment-Generating Function (mgf) - ခြေပ်ဆွဲအားဖြစ်ပေါ်စေသော ဖွင့်ဆိုချက်**
  ဒါက ကျပန်းကိန်းရှင်ရဲ့ခြေပ်စဉ်တွေ (Moments) ကို ထုတ်ပေးတဲ့ ကိရိယာတစ်ခုပါ။ M(t) = E(e^{tX}) လို့ သတ်မှတ်ပြီး ဒီကနေ Mean နဲ့ Variance ကို အလွယ်တကူ ရှာလို့ရပါတယ်။
  - E(X) = M'(0)
  - $Var(X) = M''(0) - [M'(0)]^2$

- **အရေးကြီးတဲ့ အချက်:** mgf က ဖြန့်ဖြူးမှုကို သီးသန့်သတ်မှတ်ပေးပါတယ်။ ဆိုလိုတာက mgf တူရင် ဖြန့်ဖြူးမှုလည်း တူပါတယ်။

၂.၄ Binomial Distribution (ဒွိပစ္စည်းဖြန့်ဖြူးမှု)

ဒါက အသုံးအများဆုံး Discrete Distribution တွေထဲက တစ်ခုပါ။

- **အခြေအနေ (၄) ခု**
  1. စမ်းသပ်မှု n ကြိမ် ပြုလုပ်တယ် (n က သတ်မှတ်ထားတယ်)။
  2. အကြိမ်တိုင်းဟာ အမှီအခိုကင်းတယ် (Independent)။
  3. အကြိမ်တိုင်းမှာ အောင်မြင်နိုင်ခြေ (p) က အတူတူပဲ။ (မအောင်မြင်နိုင်ခြေ q = 1-p)
  4. X က n ကြိမ်မှာ အောင်မြင်တဲ့ အကြိမ်အရေအတွက်ပါ။
- **pmf**
  f(x) = C(n,x) * p^x * (1-p)^{n-x} ဖြစ်ပြီး x = 0, 1, 2, ..., n ဖြစ်ပါတယ်။
- **Mean နဲ့ Variance**
  E(X) = np နဲ့ Var(X) = np(1-p) ဖြစ်ပါတယ်။
- **ဥပမာ - ကျောင်းသား ၁၀ ယောက်ထဲက ၆ ယောက် အောင်နိုင်ခြေ**
  ကျောင်းသားတစ်ယောက် အောင်နိုင်ခြေ p = 0.8 ဆိုရင် ကျောင်းသား ၁၀ ယောက်ထဲက အတိအကျ ၆ ယောက် အောင်နိုင်ခြေက f(6) = C(10,6) * (0.8)^6 * (0.2)^4 ဖြစ်ပါတယ်။

၂.၅ Negative Binomial Distribution (အနုတ်ဒွိပစ္စည်းဖြန့်ဖြူးမှု)

ဒါက Binomial နဲ့ ဆင်ပေမယ့် သိမ်မွေ့တဲ့ ကွာခြားချက်တစ်ခုရှိပါတယ်။

- **အဓိကကွာခြားချက်**
  Binomial က n (အကြိမ်အရေအတွက်) ကို သတ်မှတ်ပြီး
  အောင်မြင်တဲ့အကြိမ်အရေအတွက် (X) ကို ရှာတယ်။

Negative Binomial က အောင်မြင်မှုအရေအတွက် (r) ကို သတ်မှတ်ပြီး အဲဒီအောင်မြင်မှု ရဖို့ လိုအပ်တဲ့ စမ်းသပ်မှုအကြိမ်အရေအတွက် (X) ကို ရှာပါတယ်။

- **အသုံး**
  ဥပမာ - ဘတ်စကက်ဘောသမားတစ်ယောက် အောင်မြင်တဲ့အချက် ၁၀ ချက် ပစ်ဖို့ ဘယ်နှစ်ကြိမ်ပစ်ရမလဲဆိုတာ သိချင်ရင် သုံးပါတယ်။

၂.၆ Poisson Distribution (ပရွာဆန်ဖြန့်ဖြူးမှု)

ဒါက သတ်မှတ်ထားတဲ့ အချိန် ဒါမှမဟုတ် နေရာတစ်ခုမှာ ဖြစ်ရပ်တစ်ခု ဘယ်နှစ်ကြိမ် ဖြစ်တတ်သလဲဆိုတာ တွက်တဲ့ ဖြန့်ဖြူးမှုပါ။

- **အခြေအနေ (၃) ခု**
  1. ထပ်မနေတဲ့ အချိန်ပိုင်းတွေမှာ ဖြစ်ရပ်အရေအတွက်တွေဟာ အမှီအခိုကင်းတယ်။
  2. အလွန်တိုတောင်းတဲ့ အချိန်ပိုင်းလေး h မှာ ဖြစ်ရပ်တစ်ခုတိကျစွာ ဖြစ်နိုင်ခြေက λh လောက်ရှိတယ်။
  3. အလွန်တိုတောင်းတဲ့ အချိန်ပိုင်းလေးမှာ ဖြစ်ရပ် ၂ ခု ဒါမှမဟုတ် အဲဒီထက်ပိုပြီး ဖြစ်နိုင်ခြေက သိပ်ကို နည်းပါတယ်။

- **pmf**
  f(x) = (e^{-λ} * λ^x) / x! ဖြစ်ပြီး x = 0, 1, 2, ... ဖြစ်ပါတယ်။ (λ က ပျမ်းမျှဖြစ်ရပ်အရေအတွက်ပါ)

- **ထူးခြားချက်**
  Poisson Distribution ရဲ့ ထူးခြားချက်က Mean နဲ့ Variance က အတူတူပါပဲ။ E(X) = λ နဲ့ Var(X) = λ ဖြစ်ပါတယ်။

- **ဥပမာ - ဖုန်းခေါ်ဆိုမှုများ**
  တစ်နာရီကို ပျမ်းမျှ ဖုန်းခေါ်ဆိုမှု ၁၁ ကြိမ်ရှိတဲ့ ခရီးသွားအေဂျင်စီမှာ တစ်နာရီအတွင်း ဖုန်းခေါ်ဆိုမှု ၁၀ ကြိမ်ထက်ပိုဖို့ ဖြစ်တန်စွမ်းကို Poisson သုံးပြီး တွက်လို့ရပါတယ်။

- **Binomial ကို Poisson နဲ့ အနီးစပ်ဆုံးတွက်ခြင်း**
  n ကြီးပြီး p သေးရင် (np = λ ဖြစ်အောင်) Binomial ကို Poisson နဲ့ အနီးစပ်ဆုံး တွက်လို့ရပါတယ်။ ဒါက တွက်ချက်မှုကို ပိုမိုလွယ်ကူစေပါတယ်။

အချုပ်အားဖြင့် ဒီအခန်းက Discrete Random Variables ရဲ့ အခြေခံသဘောတရားတွေ၊ သူတို့ရဲ့ ဖြန့်ဖြူးမှုပုံစံတွေ (Binomial, Negative Binomial, Poisson) နဲ့ အဲဒီဖြန့်ဖြူးမှုတွေရဲ့ လက္ခဏာတွေကို ဘယ်လိုရှာသလဲဆိုတာ သင်ကြားပေးထားပါတယ်။ ဒါတွေက နောင်အခန်းတွေမှာ သင်ယူရမယ့် Statistical Inference ရဲ့ အခြေခံအုတ်မြစ် ဖြစ်ပါတယ်။

အခန်း (၃) အဆက်မပြတ် ဖြန့်ဖြူးမှုများ (Continuous Distributions)

ဒီအခန်းရဲ့အဓိက ရည်ရွယ်ချက်က **အဆက်မပြတ်သော ကျပန်းကိန်းရှင်များ (Continuous Random Variables)** နဲ့ သူတို့ရဲ့ဖြန့်ဖြူးမှု (Distribution) တွေကို လေ့လာဖို့ပါ။ Discrete Random Variables တွေမှာ အမှတ်တွေအတွက် ဖြစ်တန်စွမ်း (Probability Mass Function) ကို သုံးခဲ့ပေမယ့် Continuous Random Variables တွေမှာတော့ **Probability Density Function (pdf)** ကို သုံးပါတယ်။

ဒီမှာ သင်ယူရမှာက (၁) Continuous Random Variable ဆိုတာ ဘာလဲ၊ (၂) Exponential, Gamma, Chi-Square Distribution တွေ၊ (၃) အလွန်အရေးကြီးတဲ့ Normal Distribution နဲ့ (၄) တခြား အသုံးဝင်တဲ့ မော်ဒယ်တွေပဲ ဖြစ်ပါတယ်။

၃.၁ အဆက်မပြတ်သော ကျပန်းကိန်းရှင်များ (Random Variables of the Continuous Type)

ဒီအပိုင်းမှာ Continuous Random Variable ဆိုတာ ဘာလဲ၊ သူ့ရဲ့pdf က ဘယ်လိုအလုပ်လုပ်သလဲဆိုတာ စပြီး လေ့လာရမှာပါ။

- **Continuous Random Variable ဆိုတာ**

  တန်ဖိုးတွေကို ရေတွက်လို့မရဘဲ ကြားကာလ (Interval) အတွင်းက တန်ဖိုးတစ်ခုခု ဖြစ်နိုင်တဲ့ ကျပန်းကိန်းရှင်မျိုးပါ။ (ဥပမာ - လူတစ်ယောက်ရဲ့အရပ်၊ အလေးချိန်၊ အချိန်)။

- **Probability Density Function (pdf) - ဖြစ်တန်စွမ်းသိပ်သည်းဆဖွင့်ဆိုချက်**

  Continuous Random Variable တစ်ခုရဲ့ဖြန့်ဖြူးမှုကို ဖော်ပြတဲ့ ဖွင့်ဆိုချက်ပါ။ Discrete လိုမျိုး တန်ဖိုးတစ်ခုချင်းစီအတွက် ဖြစ်တန်စွမ်း မဟုတ်ဘဲ၊ ကြားကာလတစ်ခုအတွင်းမှာ ကျရောက်နိုင်ခြေကို ကိုယ်စားပြုပါတယ်။

- **pdf ရဲ့ဂုဏ်သတ္တိများ**

  1. $f(x) \geq 0$ (အမြဲတမ်း အနှုတ်မဖြစ်ဘူး)
  2. $\int f(x)\,dx = 1$ (စုစုပေါင်း ဧရိယာက ၁ ဖြစ်တယ်) - ဒါက ဖြစ်တန်စွမ်း အားလုံးပေါင်းက ၁ ဆိုတာနဲ့ ဆင်တူပါတယ်။

3. $P(a < X < b) = \int_a^b f(x)\,dx$ (a နဲ့ b ကြားမှာ ကျရောက်နိုင်ခြေက အဲဒီကြားက ဧရိယာပါ)

- **Cumulative Distribution Function (cdf) - အစုအပြုံဖြန့်ဖြူးမှုဖွင့်ဆိုချက်**
  $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$ ဖြစ်ပြီး pdf ရဲ့ဆက်စပ်ပစ္စည်းပါ။

- **ဥပမာ - Uniform Distribution (တူညီသောဖြန့်ဖြူးမှု)**
  X က [a, b] ကြားက တန်ဖိုးတစ်ခုခု ဖြစ်နိုင်ပြီး တန်ဖိုးအားလုံး ဖြစ်နိုင်ခြေ တူညီတယ်ဆိုရင် pdf က $f(x) = 1/(b-a)$ ဖြစ်ပါတယ်။ ဒါကို U(a,b) လို့ ခေါ်ပါတယ်။

- **မျှော်မှန်းချက်နဲ့ ကွဲလွဲမှု (Mean and Variance)**
  Discrete လိုပဲ ပေါင်းလဒ်အစား ပေါင်းစည်းခြင်း (Integral) နဲ့ တွက်ပါတယ်။

  - $E(X) = \int x f(x)\,dx$
  - $Var(X) = \int (x-\mu)^2 f(x)\,dx$

၃.၂ Exponential, Gamma နဲ့ Chi-Square Distribution များ

ဒီအပိုင်းမှာ Poisson Process နဲ့ ဆက်စပ်နေတဲ့ အရေးကြီးတဲ့ ဖြန့်ဖြူးမှု (၃) ခုကို လေ့လာရမှာပါ။

- **Exponential Distribution (ထပ်ကိန်းဖြန့်ဖြူးမှု)**

  - **အဓိပ္ပါယ်:** Poisson Process တစ်ခုမှာ ပထမဆုံး ဖြစ်ရပ် မဖြစ်ခင် စောင့်ရမယ့် အချိန်ကို ဖော်ပြတဲ့ ဖြန့်ဖြူးမှုပါ။
  - **pdf:** $f(x) = (1/\theta) * e^{-x/\theta}$ $(x \ge 0)$။ ဒီမှာ θ က ပျမ်းမျှစောင့်ဆိုင်းချိန်ပါ။
  - **Mean နဲ့ Variance:** $E(X) = \theta$ နဲ့ $Var(X) = \theta^2$
  - **ထူးခြားချက် - "မေ့ပျောက်ခြင်း" ဂုဏ်သတ္တိ (Memoryless Property):** ဒီဖြန့်ဖြူးမှုရဲ့ထူးခြားချက်က $P(X > s+t \mid X > s) = P(X > t)$ ဖြစ်တယ်။ ဆိုလိုတာက ပစ္စည်းတစ်ခု အချိန်အတော်ကြာ အလုပ်လုပ်ပြီးသားဖြစ်နေပေမယ့် နောက်ထပ် ဘယ်လောက်ကြာကြာ အလုပ်လုပ်နိုင်မလဲဆိုတဲ့ ဖြစ်တန်စွမ်းက

အသစ်စက်စက်ပစ္စည်းတစ်ခုနဲ့ အတူတူပါပဲ။ (ဒါကြောင့် ဒီမော်ဒယ်က စက်ပစ္စည်းတွေရဲ့ သက်တမ်းအတွက် သိပ်မသင့်တော်ပါဘူး)

- **Gamma Distribution (ဂမ်မာဖြန့်ဖြူးမှု)**
    - **အဓိပ္ပါယ်:** Poisson Process တစ်ခုမှာ α ကြိမ်မြောက် ဖြစ်ရပ် မဖြစ်ခင် စောင့်ရမယ့် အချိန်ကို ဖော်ပြတဲ့ ဖြန့်ဖြူးမှုပါ။
    - **pdf:** f(x) = [1 / (Γ(α) * θ^α)] * x^{α-1} * e^{-x/θ} (x ≥ 0)။
    - **Mean နဲ့ Variance:** E(X) = αθ နဲ့ Var(X) = αθ²
    - **Gamma Function:** Γ(α) = ∫_0^∞ y^{α-1} e^{-y} dy ဖြစ်ပြီး α က ကိန်းပြည့်ဆိုရင် Γ(α) = (α-1)! ဖြစ်ပါတယ်။ (ဒါကြောင့် Gamma Function ကို Generalized Factorial လို့ ခေါ်တာပါ)
- **Chi-Square Distribution (ခိုင်နှစ်ထပ်ဖြန့်ဖြူးမှု)**
    - **အဓိပ္ပါယ်:** Gamma Distribution ရဲ့ အထူးအခြေအနေတစ်ခုပါ။ α = r/2 နဲ့ θ = 2 ဆိုရင် Chi-Square Distribution ဖြစ်သွားပါတယ်။ r ကို ဒီဂရီလွတ်လပ်မှု (Degrees of Freedom) လို့ ခေါ်ပါတယ်။
    - **အသုံး:** စာရင်းအင်းမှာ အလွန်အရေးပါပြီး Variance ကို ခန့်မှန်းခြင်း၊ Goodness-of-Fit စစ်ဆေးခြင်းတွေမှာ သုံးပါတယ်။

၃.၃ Normal Distribution (ပုံမှန်ဖြန့်ဖြူးမှု)

ဒါက စာရင်းအင်းမှာ အကျော်ကြားဆုံးနဲ့ အသုံးအများဆုံး ဖြန့်ဖြူးမှုပါ။

- **အဓိပ္ပါယ်**

  X ရဲ့ pdf က f(x) = [1 / (σ√(2π))] * e^{-(x-μ)² / (2σ²)} ဖြစ်ပြီး ဒါကို N(μ, σ²) လို့ ရေးပါတယ်။

- **ပုံသဏ္ဌာန်**
  ခေါင်းလောင်းပုံ (Bell-shaped) ဖြစ်ပြီး ပျမ်းမျှ μ မှာ အမြင့်ဆုံးဖြစ်ကာ နှစ်ဖက်စလုံး အချိုးကျ (Symmetric) ပါတယ်။

- **Standard Normal Distribution (စံပုံမှန်ဖြန့်ဖြူးမှု)**
  μ = 0 နဲ့ σ = 1 ဖြစ်တဲ့ Normal Distribution ကို Z ~ N(0,1) လို့ ခေါ် ပြီး ဒီဇယားကို သုံးပြီး ဖြစ်တန်စွမ်းအားလုံးကို တွက်လို့ရပါတယ်။

- **အရေးကြီးတဲ့ သီအိုရမ်**
  $X \sim N(\mu, \sigma^2)$ ဆိုရင် $Z = (X - \mu) / \sigma$ က N(0,1) ဖြစ်ပါတယ်။ (ဒါက Standardization လုပ်တာပါ)

- **ဥပမာ - IQ Score**
  IQ Score တွေက ပျမ်းမျှ ၁၀၀ နဲ့ စံသွေဖည်မှု ၁၅ နဲ့ Normal Distribution အတိုင်း ဖြန့်ဖြူးတယ်ဆိုရင် IQ ၁၁၅ ထက်ပိုတဲ့သူရဲ့ ရာခိုင်နှုန်းကို Standard Normal Table သုံးပြီး တွက်လို့ရပါတယ်။

- **Chi-Square နဲ့ ဆက်စပ်မှု**
  $Z \sim N(0,1)$ ဆိုရင် $Z^2 \sim \chi^2(1)$ ဖြစ်ပါတယ်။

၃.၄ အပိုမော်ဒယ်များ (Additional Models)

ဒီအပိုင်းမှာ သက်တမ်းခွဲခြမ်းစိတ်ဖြာခြင်း (Life Testing) မှာ အသုံးများတဲ့ မော်ဒယ်တွေကို လေ့လာရမှာပါ။

- **Failure Rate (ချို့ယွင်းမှုနှုန်း)**
  အချိန် t မှာ ပစ္စည်းတစ်ခု ချို့ယွင်းနိုင်ခြေကို λ(t) နဲ့ ဖော်ပြပါတယ်။ Exponential Distribution မှာ λ(t) က ကိန်းသေဖြစ်ပြီး (အသစ်နဲ့အဟောင်း အတူတူ)၊ တခြားမော်ဒယ်တွေမှာ အချိန်နဲ့အမျှ တိုးလာပါတယ်။

- **Weibull Distribution (ဝိုင်ဘောဖြန့်ဖြူးမှု)**
  စက်မှုလုပ်ငန်းတွေမှာ ပစ္စည်းတွေရဲ့သက်တမ်းကို ဖော်ပြဖို့ အသုံးများတဲ့ မော်ဒယ်ပါ။ failure rate က λ(t) = (α/β^α) * t^{α-1} ဖြစ်ပြီး α က ၁ ထက်ကြီးရင် အချိန်နဲ့အမျှ တိုးလာပါတယ်။

- **Gompertz Law (ဂွမ်ပါ့ဇ်နိယာမ)**
  လူသားတွေရဲ့သေဆုံးမှုနှုန်းကို ဖော်ပြတဲ့ မော်ဒယ်ပါ။ failure rate က λ(t) = a * e^{bt} ဖြစ်ပြီး အသက်ကြီးလာတာနဲ့အမျှ သေဆုံးနိုင်ခြေ ပိုများလာတာကို ဖော်ပြတယ်။ (ဥပမာ - အသက် ၂၅ နှစ်နောက်ပိုင်း သေဆုံးမှုနှုန်း ၁၀% ခန့် တိုးလာတယ်)

အချုပ်အားဖြင့် ဒီအခန်းက Continuous Random Variables တွေရဲ့အခြေခံသဘောတရားတွေ၊ သူတို့ရဲ့ pdf နဲ့ cdf ဆိုတာတွေ၊ အရေးကြီးတဲ့ ဖြန့်ဖြူးမှုပုံစံတွေ (Exponential, Gamma, Chi-Square, Normal) နဲ့ လက်တွေ့ဘဝမှာ အသုံးများတဲ့ Weibull နဲ့ Gompertz မော်ဒယ်တွေအကြောင်း သင်ကြားပေးထားပါတယ်။ Normal Distribution ရဲ့အရေးပါမှုက နောက်အခန်းတွေမှာ သင်ယူရမယ့် Central Limit Theorem ကြောင့် ပိုပြီး ထင်ရှားလာမှာပါ။

အခန်း (၄) နှစ်ခုမြောက် ဖြန့်ဖြူးမှုများ (Bivariate Distributions)

ဒီအခန်းရဲ့ အဓိက ရည်ရွယ်ချက်က **ကျပန်းကိန်းရှင်နှစ်ခု (Two Random Variables) ပေါင်းစပ်ထားတဲ့ ဖြန့်ဖြူးမှုတွေ (Bivariate Distributions)** ကို လေ့လာဖို့ပါ။ ဒီအခန်းမှာ သင်ယူရမှာက (၁) Bivariate Distributions ဆိုတာ ဘာလဲ၊ (၂) Correlation Coefficient (ဆက်စပ်မှုကိန်း)၊ (၃) Conditional Distributions (အခြေအနေအရ ဖြန့်ဖြူးမှု) နဲ့ (၄) Bivariate Normal Distribution တို့ပဲ ဖြစ်ပါတယ်။

ဒီအခန်းရဲ့ ရည်ရွယ်ချက်က ကိန်းရှင်နှစ်ခုကြား ဆက်စပ်မှုကို နားလည်ဖို့ပါ။ ဥပမာ - ကျောင်းသားတစ်ယောက်ရဲ့ အရပ်နဲ့ ကိုယ်အလေးချိန်၊ ဒါမှမဟုတ် သူ့ရဲ့ SAT ရမှတ်နဲ့ ပထမနှစ် စာမေးပွဲရမှတ်တို့ကြား ဆက်စပ်မှုကို လေ့လာတာမျိုးပါ။

၄.၁ အပိုင်းအခြားသတ်မှတ်နိုင်သော နှစ်ခုမြောက်ဖြန့်ဖြူးမှုများ (Bivariate Distributions of the Discrete Type)

ဒီအပိုင်းမှာ ကျပန်းကိန်းရှင်နှစ်ခု (X နဲ့ Y) ရဲ့ဖြစ်တန်စွမ်းကို ဘယ်လို အတူတကွ သတ်မှတ်သလဲဆိုတာ လေ့လာရမှာပါ။

- **Joint Probability Mass Function (Joint pmf) - ပေါင်းစပ်ဖြစ်တန်စွမ်းဩဇာထုဖွင့်ဆိုချက်**

  ဒါက X နဲ့ Y နှစ်ခုလုံးရဲ့တန်ဖိုးတွေ အတူတူ ဖြစ်နိုင်ခြေကို သတ်မှတ်တဲ့ ဖွင့်ဆိုချက်ပါ။ f(x,y) = P(X=x, Y=y) လို့ ရေးပါတယ်။

  - ဂုဏ်သတ္တိ (၃) ခုရှိတယ်။

    ၁။ $f(x,y) \geq 0$

    ၂။ $\Sigma \Sigma f(x,y) = 1$ (အားလုံးပေါင်းက ၁)

    ၃။ $P[(X,Y) \in A] = \Sigma_{(x,y)\in A} f(x,y)$

- **Marginal Probability Mass Function (Marginal pmf) - အနားသတ်ဖြစ်တန်စွမ်းဩဇာထုဖွင့်ဆိုချက်**

ဒါက X (သို့မဟုတ် Y) တစ်ခုတည်းရဲ့ဖြစ်တန်စွမ်းကို ပြောတာပါ။ Joint pmf ကနေ တစ်ခုကို ပေါင်းလိုက်ရင် ရပါတယ်။

- X အတွက်: $f_X(x) = \Sigma_y f(x,y)$
- Y အတွက်: $f_Y(y) = \Sigma_x f(x,y)$

- **လွတ်လပ်သော ကျပန်းကိန်းရှင်များ (Independent Random Variables)**
  X နဲ့ Y က လွတ်လပ်တယ်ဆိုရင် $f(x,y) = f_X(x) * f_Y(y)$ ဖြစ်ရပါမယ်။ ဆိုလိုတာက X ရဲ့တန်ဖိုးက Y ရဲ့ဖြစ်တန်စွမ်းကို သက်ရောက်မှုမရှိဘူးပေါ့။

- **ဥပမာ - အန်စာတုံးနှစ်ခု**
  အန်စာတုံးနှစ်ခု လှိမ့်တယ်။ X က သေးငယ်တဲ့အရေအတွက်၊ Y က ကြီးတဲ့အရေအတွက်ဆိုရင် Joint pmf ကို ဇယားတစ်ခုအနေနဲ့ ဖော်ပြနိုင်ပါတယ်။
  - ဥပမာ - $P(X=2, Y=3) = 2/36$ (ဘာကြောင့်လဲဆိုတော့ (2,3) နဲ့ (3,2) ဆိုတဲ့ နည်းလမ်း ၂ ခုရှိလို့)။
  - $P(X=2)$ ကို ရှာချင်ရင် Y အားလုံးကို ပေါင်းလိုက်ရုံပါပဲ။

၄.၂ ဆက်စပ်မှုကိန်း (The Correlation Coefficient)

ဒီအပိုင်းမှာ ကိန်းရှင်နှစ်ခုကြား ဆက်စပ်မှု အတိုင်းအတာကို လေ့လာရမှာပါ။

- **Covariance (ပေါင်းစပ်ကွဲလွဲမှု)**
  X နဲ့ Y က သူတို့ရဲ့ပျမ်းမျှကနေ ဘယ်လိုအတူတကွ ကွဲလွဲသလဲဆိုတဲ့ အတိုင်းအတာပါ။
  - $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X)E(Y)$

- **Correlation Coefficient (ဆက်စပ်မှုကိန်း)**
  Covariance ကို ကိန်းရှင်တစ်ခုချင်းစီရဲ့စံသွေဖည်မှုနဲ့ စားထားတဲ့ တန်ဖိုးပါ။ ဒါက -၁ နဲ့ +၁ ကြားမှာ ရှိပါတယ်။
  - $\rho = Cov(X,Y) / (\sigma_X * \sigma_Y)$

- **ဆက်စပ်မှုကိန်းရဲ့အဓိပ္ပါယ်**
    - ρ က +၁ နီးရင် X ကြီးလာတာနဲ့ Y လည်း ကြီးလာတယ် (တိုက်ရိုက်ဆက်စပ်မှု)။
    - ρ က -၁ နီးရင် X ကြီးလာတာနဲ့ Y ငယ်လာတယ် (ပြောင်းပြန်ဆက်စပ်မှု)။
    - ρ က ၀ နီးရင် X နဲ့ Y ကြားမှာ မျဉ်းဖြောင့်ဆက်စပ်မှု မရှိဘူး (ဒါပေမယ့် တခြားပုံစံ ဆက်စပ်မှုတော့ ရှိနိုင်တယ်)။
- **ဥပမာ - အရပ်နဲ့ ကိုယ်အလေးချိန်**
  ယေဘုယျအားဖြင့် အရပ်ရှည်လေလေ ကိုယ်အလေးချိန်များလေလေ ဖြစ်တဲ့အတွက် ဒီနှစ်ခုကြား Correlation Coefficient က အပြုသဘောဆောင်ပါတယ်။

၄.၃ အခြေအနေအရ ဖြန့်ဖြူးမှုများ (Conditional Distributions)

ဒီအပိုင်းမှာ X ရဲ့တန်ဖိုးကို သိပြီးသားအခါ Y ရဲ့ဖြန့်ဖြူးမှု ဘယ်လိုရှိသလဲဆိုတာ လေ့လာရမှာပါ။

- **Conditional pmf (အခြေအနေအရ ဖြစ်တန်စွမ်းဩဇာဆိုချက်)**
  X = x ဆိုတဲ့ အခြေအနေအောက်မှာ Y ရဲ့ဖြစ်တန်စွမ်းကို အောက်ပါအတိုင်း သတ်မှတ်ပါတယ်။
    - $h(y|x) = P(Y=y \mid X=x) = f(x,y) / f_X(x)$
- **Conditional Mean နဲ့ Conditional Variance**
    - $E(Y \mid X=x) = \Sigma_y [y * h(y|x)]$ (X = x ဖြစ်တဲ့အခါ Y ရဲ့ပျမ်းမျှ)
    - $Var(Y \mid X=x) = \Sigma_y [(y - E(Y|x))^2 * h(y|x)]$
- **ဥပမာ - သွေးပေါင်ချိန်နဲ့ နှလုံးခုန်မမှန်ခြင်း**
  လူနာတစ်ယောက်ရဲ့သွေးပေါင်ချိန် (X) ကို သိထားပြီးသားဆိုရင် သူမှာ နှလုံးခုန်မမှန်ခြင်း (Y) ရှိမရှိရဲ့ဖြစ်တန်စွမ်းကို Conditional Distribution နဲ့ တွက်လို့ရပါတယ်။

၄.၄ အဆက်မပြတ်သော နှစ်ခုမြောက်ဖြန့်ဖြူးမှုများ (Bivariate Distributions of the Continuous Type)

Discrete အတွက် သင်ယူခဲ့တဲ့ အယူအဆတွေကို Continuous အတွက် သက်ဆိုင်အောင် ပြောင်းလဲသုံးစွဲတာပါ။

- **Joint Probability Density Function (Joint pdf) - ပေါင်းစပ်ဖြစ်တန်စွမ်းသိပ်သည်းဆဖွင့်ဆိုချက်**

  Discrete မှာ ပေါင်းလဒ် (Sum) သုံးခဲ့ရင် Continuous မှာ ပေါင်းစည်းခြင်း (Integral) ကို သုံးပါတယ်။

  - $\int\int f(x,y)\,dx\,dy = 1$
  - $P[(X,Y) \in A] = \int\int_A f(x,y)\,dx\,dy$

- **Marginal pdf**
  - $f_X(x) = \int f(x,y)\,dy$
  - $f_Y(y) = \int f(x,y)\,dx$
- **Conditional pdf**
  - $h(y|x) = f(x,y) / f_X(x)$
- **ဥပမာ - ဧရိယာတစ်ခုအတွင်း ကျရောက်နိုင်ခြေ**

  $f(x,y) = 2$ ဖြစ်ပြီး $0 \le x \le y \le 1$ ဆိုတဲ့ ဧရိယာထဲမှာ ရှိတယ်ဆိုရင် $P(0 \le X \le 1/2, 0 \le Y \le 1/2)$ ကို Double Integral သုံးပြီး တွက်လို့ရပါတယ်။

၄.၅ နှစ်ခုမြောက်ပုံမှန်ဖြန့်ဖြူးမှု (The Bivariate Normal Distribution)

ဒါက နှစ်ခုမြောက်ဖြန့်ဖြူးမှုတွေထဲက အရေးအကြီးဆုံးနဲ့ အသုံးအများဆုံး ဖြန့်ဖြူးမှုပါ။

- **အဓိပ္ပါယ်**

  X နဲ့ Y နှစ်ခုလုံးက Normal Distribution အတိုင်း ဖြန့်ဖြူးပြီး သူတို့ကြား ဆက်စပ်မှုရှိတယ်ဆိုရင် ဒီဖြန့်ဖြူးမှုကို Bivariate Normal Distribution လို့ ခေါ်ပါတယ်။ ဒါကို သတ်မှတ်ဖို့ ပါရာမီတာ (၅) ခု လိုအပ်ပါတယ်။ $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$

- **အရေးကြီးတဲ့ ဂုဏ်သတ္တိများ**

    1. **Conditional Distribution:** X = x ဆိုတဲ့ အခြေအနေအောက်မှာ Y ရဲ့ဖြန့်ဖြူးမှုက Normal ဖြစ်တယ်။

        - $E(Y \mid X=x) = \mu_Y + \rho * (\sigma_Y/\sigma_X) * (x - \mu_X)$ (ဒါက မျဉ်းဖြောင့်ဖြစ်တယ်)
        - $Var(Y \mid X=x) = \sigma_Y^2 * (1 - \rho^2)$

    2. **လွတ်လပ်မှု:** ဒီဖြန့်ဖြူးမှုမှာ $\rho = 0$ ဆိုရင် X နဲ့ Y က လွတ်လပ်ပါတယ်။ (တခြားဖြန့်ဖြူးမှုတွေမှာ ဒါမမှန်နိုင်ဘူး)

    3. **ပုံသဏ္ဌာန်:** ဒီဖြန့်ဖြူးမှုရဲ့ဂရပ်က တောင်ကုန်းပုံစံဖြစ်ပြီး မြေပုံပေါ်မှာ ဘဲဥပုံ (Ellipse) အနေနဲ့ မြင်ရပါတယ်။

- **ဥပမာ - GPA နဲ့ SAT**

    ကျောင်းသားတစ်ယောက်ရဲ့ High School GPA (X) နဲ့ SAT Score (Y) က Bivariate Normal Distribution နဲ့ ဖြန့်ဖြူးတယ်ဆိုရင် GPA ၃.၅ ရှိတဲ့ကျောင်းသားရဲ့ SAT Score ပျမ်းမျှကို Conditional Mean ဖော်မြူလာနဲ့ တွက်လို့ရပါတယ်။

အချုပ်အားဖြင့် ဒီအခန်းက ကျပန်းကိန်းရှင်နှစ်ခုပါတဲ့ ဖြန့်ဖြူးမှုတွေရဲ့အခြေခံသဘောတရားတွေ၊ သူတို့ကြားက ဆက်စပ်မှုကို တိုင်းတာတဲ့ Correlation Coefficient၊ အခြေအနေအရ ဖြန့်ဖြူးမှုတွေနဲ့ အရေးအကြီးဆုံးဖြစ်တဲ့ Bivariate Normal Distribution အကြောင်း သင်ကြားပေးထားပါတယ်။ ဒါတွေက စာရင်းအင်းမှာ ကိန်းရှင်နှစ်ခုကြား ဆက်စပ်မှုကို နားလည်ဖို့ အခြေခံအုတ်မြစ် ဖြစ်ပါတယ်။

အခန်း (၅) ကျပန်းကိန်းရှင်များ၏ လုပ်ဆောင်ချက်များ၏ ဖြန့်ဖြူးမှု

ဒီအခန်းရဲ့အဓိက ရည်ရွယ်ချက်က **ကျပန်းကိန်းရှင်တွေရဲ့လုပ်ဆောင်ချက်တွေ (Functions of Random Variables)** ရဲ့ဖြန့်ဖြူးမှုကို ဘယ်လိုရှာသလဲဆိုတာ လေ့လာဖို့ပါ။ ဒီအခန်းမှာ သင်ယူရမှာက (၁) ကျပန်းကိန်းရှင်တစ်ခုရဲ့လုပ်ဆောင်ချက်၊ (၂) ကျပန်းကိန်းရှင်နှစ်ခုရဲ့ အသွင်ပြောင်းလုပ်ဆောင်ချက်၊ (၃) ကျပန်းကိန်းရှင်များစွာ၊ (၄) Moment-Generating Function (mgf) နည်းပညာ၊ (၅) Normal Distribution နဲ့ ဆက်စပ်တဲ့ ကျပန်းလုပ်ဆောင်ချက်များ၊ (၆) Central Limit Theorem (CLT)၊ (၇) Discrete Distribution တွေအတွက် အနီးစပ်ဆုံး တွက်နည်းများ၊ (၈) Chebyshev's Inequality နဲ့ (၉) Limiting Moment-Generating Functions တို့ပဲ ဖြစ်ပါတယ်။

ဒီအခန်းရဲ့အထူးအရေးပါတဲ့ အပိုင်းကတော့ **Central Limit Theorem (CLT)** ပါ။ ဒါက စာရင်းအင်းရဲ့အခြေခံအုတ်မြစ်တစ်ခုလို့ တင်စားလို့ရပါတယ်။

၅.၁ ကျပန်းကိန်းရှင်တစ်ခု၏ လုပ်ဆောင်ချက်များ (Functions of One Random Variable)

ဒီအပိုင်းမှာ ကျပန်းကိန်းရှင် X ရဲ့လုပ်ဆောင်ချက်တစ်ခုဖြစ်တဲ့ Y = u(X) ရဲ့ဖြန့်ဖြူးမှုကို ဘယ်လိုရှာသလဲဆိုတာ လေ့လာရမှာပါ။

- **Distribution Function Technique (ဖြန့်ဖြူးမှုဖွင့်ဆိုချက်နည်းပညာ)**
  ဒါက Y ရဲ့ cdf ကို အရင်ရှာပြီး နောက်မှ pdf ကို ဆင်းသက်ယူတဲ့ နည်းလမ်းပါ။
  - G(y) = P(Y ≤ y) = P(u(X) ≤ y)
  - ဒီကနေ g(y) = G'(y) လို့ ရှာလို့ရပါတယ်။
- **Change-of-Variable Technique (ကိန်းရှင်ပြောင်းလဲခြင်းနည်းပညာ)**
  Y = u(X) က တစ်ခုတည်းသော ပြောင်းပြန်လုပ်ဆောင်ချက် (One-to-One) ရှိရင် ဒီနည်းက ပိုမြန်ပါတယ်။
  - g(y) = f(v(y)) * |v'(y)| (ဒီမှာ x = v(y) က ပြောင်းပြန်လုပ်ဆောင်ချက်ပါ)

- **ဥပမာ - အန်စာတုံးတစ်ခုရဲ့နှစ်ထပ်ကိန်း**
  X က U(0,1) ဖြစ်ပြီး $Y = X^2$ ဆိုရင် Y ရဲ့ pdf က $g(y) = 1/(2\sqrt{y})$ ဖြစ်ပါတယ်။

၅.၂ ကျပန်းကိန်းရှင်နှစ်ခု၏ အသွင်ပြောင်းလုပ်ဆောင်ချက်များ (Transformations of Two Random Variables)

ဒီအပိုင်းမှာ ကျပန်းကိန်းရှင်နှစ်ခု (X_1 နဲ့ X_2) ရဲ့လုပ်ဆောင်ချက်တွေ (Y_1 နဲ့ Y_2) ရဲ့ဖြန့်ဖြူးမှုကို ဘယ်လိုရှာသလဲဆိုတာ လေ့လာရမှာပါ။

- **Change-of-Variable Technique (နှစ်ခုအတွက် ကိန်းရှင်ပြောင်းလဲခြင်းနည်းပညာ)**
  ဒီမှာ Jacobian ဆိုတဲ့ အယူအဆအသစ် ပါလာပါတယ်။
  - $g(y_1, y_2) = f(v_1(y_1,y_2), v_2(y_1,y_2)) * |J|$
  - J က Jacobian Determinant ဖြစ်ပြီး $|J| = |\partial(x_1,x_2) / \partial(y_1,y_2)|$ လို့ တွက်ပါတယ်။
- **အရေးကြီးတဲ့ ဖြန့်ဖြူးမှု (၂) ခု**
  1. **Beta Distribution:** X_1 နဲ့ X_2 က Gamma Distribution တွေဖြစ်ပြီး $Y_1 = X_1/(X_1+X_2)$ ဆိုရင် Y_1 က Beta Distribution ဖြစ်သွားပါတယ်။
  2. **F Distribution:** U နဲ့ V က Chi-Square Distribution တွေဖြစ်ပြီး $F = (U/r_1) / (V/r_2)$ ဆိုရင် F က F Distribution ဖြစ်သွားပါတယ်။ (ဒါက ANOVA မှာ အရမ်းအရေးပါတယ်)

၅.၃ ကျပန်းကိန်းရှင်များစွာ (Several Random Variables)

ဒီအပိုင်းမှာ ကျပန်းကိန်းရှင် n ခုအထိ ချဲ့ထွင်ပြီး သူတို့ရဲ့ပေါင်းစပ်ဖြန့်ဖြူးမှုကို လေ့လာရမှာပါ။

- **လွတ်လပ်သော ကျပန်းကိန်းရှင်များ (Independent Random Variables)**
  $X_1, X_2, \ldots, Xn$ တွေ လွတ်လပ်တယ်ဆိုရင် သူတို့ရဲ့ Joint pdf/pmf က တစ်ခုချင်းစီရဲ့ pdf/pmf တွေ မြှောက်လဒ်ဖြစ်ပါတယ်။
  - $f(x_1, x_2, \ldots, xn) = f_1(x_1) * f_2(x_2) * \ldots * f_n(xn)$

- **Random Sample (ကျပန်းနမူနာ)**
  ဒါက အတူတူဖြန့်ဖြူးမှုရှိပြီး လွတ်လပ်တဲ့ ကျပန်းကိန်းရှင်တွေရဲ့ စုစည်းမှုပါ။
    - $X_1, X_2, \ldots, Xn$ က ပျမ်းမျှ μ နဲ့ Variance σ^2 ရှိတဲ့ ဖြန့်ဖြူးမှုတစ်ခုကနေ Random Sample ဆိုရင် နမူနာပျမ်းမျှ (Sample Mean) $\bar{X}$ ရဲ့ပျမ်းမျှက μ ဖြစ်ပြီး Variance က σ^2/n ဖြစ်ပါတယ်။

၅.၄ Moment-Generating Function (mgf) နည်းပညာ

ဒီအပိုင်းမှာ mgf ကို သုံးပြီး ကျပန်းကိန်းရှင်တွေရဲ့ပေါင်းလဒ် (Sum) ရဲ့ဖြန့်ဖြူးမှုကို ဘယ်လိုရှာသလဲဆိုတာ လေ့လာရမှာပါ။

- **အဓိက သီအိုရမ်**
  $X_1, X_2, \ldots, Xn$ က လွတ်လပ်တယ်ဆိုရင် $Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$ ရဲ့ mgf က $M_Y(t) = M_\{X_1\}(a_1t) * M_\{X_2\}(a_2t) * \ldots * M_\{X_n\}(a_nt)$ ဖြစ်ပါတယ်။

- **ဥပမာ - Binomial Distribution**
  $X_1, X_2, \ldots, Xn$ က Bernoulli(p) ကနေ Random Sample ဆိုရင် $Y = \Sigma X_i$ ရဲ့ mgf က $M_Y(t) = (q + pe\^t)\^n$ ဖြစ်ပြီး ဒါက Binomial(n,p) ရဲ့ mgf ပါ။

၅.၅ Normal Distribution နဲ့ ဆက်စပ်တဲ့ ကျပန်းလုပ်ဆောင်ချက်များ

ဒီအပိုင်းမှာ Normal Distribution ကနေ ယူတဲ့ Sample Mean နဲ့ Sample Variance ရဲ့ ဖြန့်ဖြူးမှုကို လေ့လာရမှာပါ။

- **Sample Mean ($\bar{X}$)**
  $X_1, X_2, \ldots, Xn$ က $N(\mu, \sigma^2)$ ကနေ Random Sample ဆိုရင် $\bar{X} \sim N(\mu, \sigma^2/n)$ ဖြစ်ပါတယ်။

- **Sample Variance (S^2)**
  $(n-1)S^2 / \sigma^2 \sim \chi^2(n-1)$ ဖြစ်ပြီး $\bar{X}$ နဲ့ S^2 က လွတ်လပ်ပါတယ်။ (ဒါက အရမ်းအရေးကြီးတဲ့ ရလဒ်ပါ)

- **Student's t Distribution**
  $T = (\bar{X} - \mu) / (S/\sqrt{n})$ က $t(n-1)$ ဖြန့်ဖြူးမှုအတိုင်း ဖြစ်ပါတယ်။ ဒါက μ ကို မသိရင် Confidence Interval တွက်ဖို့ သုံးပါတယ်။

၅.၆ Central Limit Theorem (ဗဟိုကန့်သတ်သီအိုရမ် - CLT)

ဒါက ဒီအခန်းရဲ့ အရေးအကြီးဆုံး အပိုင်းပါ။

- **သီအိုရမ်**
  $X_1, X_2, \ldots, Xn$ က ပျမ်းမျှ μ နဲ့ Variance σ^2 ရှိတဲ့ **မည်သည့်ဖြန့်ဖြူးမှုမှမဆို** Random Sample ဆိုရင် n ကြီးလာတာနဲ့အမျှ $Z = (\bar{X} - \mu) / (\sigma/\sqrt{n})$ ရဲ့ဖြန့်ဖြူးမှုက Standard Normal Distribution N(0,1) ဆီကို ချဉ်းကပ်လာပါတယ်။

- **ဘာကြောင့် အရေးကြီးတာလဲ**
  ဒါက ဘယ်လိုဖြန့်ဖြူးမှုမျိုးမဆို (ပုံမှန်မဟုတ်ရင်တောင်) နမူနာအရွယ်အစား လုံလောက်ရင် နမူနာပျမ်းမျှရဲ့ဖြန့်ဖြူးမှုက ပုံမှန်ဖြစ်သွားတယ်လို့ ဆိုပါတယ်။ ဒါက စာရင်းအင်းခန့်မှန်းချက်တွေအတွက် အရမ်းအသုံးဝင်ပါတယ်။

- **ဥပမာ - အန်စာတုံး ၃၀ ရဲ့ပျမ်းမျှ**
  အန်စာတုံး ၃၀ ရဲ့ပျမ်းမျှအရေအတွက်က ပုံမှန်မဟုတ်တဲ့ ဖြန့်ဖြူးမှုကနေ လာပေမယ့် နမူနာအရွယ်အစား ၃၀ လောက်ဆိုရင် ဒီပျမ်းမျှရဲ့ဖြန့်ဖြူးမှုက ပုံမှန်နီးပါး ဖြစ်သွားပါတယ်။

၅.၇ Discrete Distribution တွေအတွက် အနီးစပ်ဆုံး တွက်နည်းများ

ဒီအပိုင်းမှာ Binomial နဲ့ Poisson လိုမျိုး Discrete Distribution တွေကို Normal Distribution နဲ့ ဘယ်လို အနီးစပ်ဆုံး တွက်သလဲဆိုတာ လေ့လာရမှာပါ။

- **Continuity Correction (အဆက်ပြတ်မှုအတွက် ပြင်ဆင်ချက်)**
  Discrete ကို Continuous နဲ့ အနီးစပ်ဆုံးတွက်တဲ့အခါ အမှတ်တစ်ခုအတွက်

ဖြစ်တန်စွမ်းကို အကွာအဝေးတစ်ခုအနေနဲ့ သတ်မှတ်ပါတယ်။ ဥပမာ - P(X = k) ကို P(k - 0.5 < X < k + 0.5) အနေနဲ့ တွက်ပါတယ်။

- **ဥပမာ - Binomial(100, 0.5)**

  X က Binomial(100, 0.5) ဆိုရင် P(40 ≤ X ≤ 60) ကို Normal နဲ့ အနီးစပ်ဆုံး တွက်လို့ရပါတယ်။

၅.၈ Chebyshev's Inequality နဲ့ Convergence in Probability

- **Chebyshev's Inequality**
  ဒါက ဖြန့်ဖြူးမှုပုံစံ မသိရင်တောင် ကျပန်းကိန်းရှင်တစ်ခုရဲ့ တန်ဖိုးက သူ့ရဲ့ပျမ်းမျှကနေ ဘယ်လောက်ကွာသွားနိုင်တယ်ဆိုတဲ့ အကွာအဝေးကို ပေးပါတယ်။

  - $P(|X - \mu| \geq k\sigma) \leq 1/k^2$
  - ဥပမာ - ပျမ်းမျှကနေ ၂ ဆ စံသွေဖည်မှုထက် ပိုကွာဖို့ ဖြစ်တန်စွမ်းက ၁/၄ ထက် မပိုဘူး။

- **Law of Large Numbers (ကြီးမားသောကိန်းဂဏန်းများ၏ နိယာမ)**
  n ကြိမ်လောတာနဲ့အမျှ နမူနာပျမ်းမျှ $\bar{X}$ က လူဦးရေပျမ်းမျှ μ ဆီကို ချဉ်းကပ်လာပါတယ်။ ဒါကို "$\bar{X}$ က μ ဆီကို ဖြစ်တန်စွမ်းအရ ဆုံစည်းတယ်" (Converges in Probability) လို့ ခေါ်ပါတယ်။

၅.၉ Limiting Moment-Generating Functions (ကန့်သတ် mgf များ)

ဒီအပိုင်းမှာ mgf တွေရဲ့ ကန့်သတ်တန်ဖိုးကို သုံးပြီး ဖြန့်ဖြူးမှုတွေရဲ့ အနီးစပ်ဆုံးပုံစံကို ရှာတဲ့နည်းကို လေ့လာရမှာပါ။

- **အဓိက အယူအဆ**
  mgf တွေရဲ့ ကန့်သတ်တန်ဖိုးက တခြား ဖြန့်ဖြူးမှုတစ်ခုရဲ့ mgf နဲ့ တူသွားရင် အဲဒီဖြန့်ဖြူးမှုက အနီးစပ်ဆုံး ဖြန့်ဖြူးမှု ဖြစ်ပါတယ်။

- **ဥပမာ - Poisson ကို Binomial အနီးစပ်ဆုံး**
  n ကြီးပြီး p သေးတဲ့အခါ Binomial ရဲ့ mgf က Poisson ရဲ့ mgf ဆီကို
  ချဉ်းကပ်လာပါတယ်။ ဒါက Poisson က Binomial ရဲ့ အနီးစပ်ဆုံးပုံစံ ဖြစ်တယ်ဆိုတာ
  ပြသချက်ပါ။

အချုပ်အားဖြင့် ဒီအခန်းက ကျပန်းကိန်းရှင်တွေရဲ့ လုပ်ဆောင်ချက်တွေရဲ့ဖြန့်ဖြူးမှုကို ရှာတဲ့ နည်းလမ်းတွေ (Distribution Function Technique, Change of Variable, mgf Technique)၊ အရေးကြီးတဲ့ ဖြန့်ဖြူးမှုတွေ (Beta, F, t)၊ **Central Limit Theorem** နဲ့ သူ့ရဲ့အသုံးချပုံတွေကို သင်ကြားပေးထားပါတယ်။ CLT က စာရင်းအင်းရဲ့ အခြေခံအကျဆုံး သီအိုရီတစ်ခုဖြစ်လို့ ဒီအခန်းကို ကောင်းကောင်း နားလည်ဖို့ အရမ်းအရေးကြီးပါတယ်။

အခန်း (၆) အမှတ်ခန့်မှန်းခြင်း (Point Estimation)

ဒီအခန်းရဲ့အဓိက ရည်ရွယ်ချက်က **ကိန်းဂဏန်းအချက်အလက်တွေကနေ လူဦးရေရဲ့ သွင်ပြင်လက္ခဏာ (Parameter) တွေကို ဘယ်လို ခန့်မှန်းမလဲ (Point Estimation)** ဆိုတာကို လေ့လာဖို့ပါ။ ဒီအခန်းမှာ သင်ယူရမှာက (၁) Descriptive Statistics (ဖော်ပြချက်စာရင်းဇယား)၊ (၂) Exploratory Data Analysis (စူးစမ်းရှာဖွေရေး ဒေတာခွဲခြမ်းစိတ်ဖြာခြင်း)၊ (၃) Order Statistics (အစီအစဉ်စာရင်းဇယား)၊ (၄) Maximum Likelihood Estimation (MLE)၊ (၅) Regression (ဆုတ်ယုတ်ခွဲခြမ်းစိတ်ဖြာခြင်း)၊ (၆) Sufficient Statistics (လုံလောက်သော စာရင်းဇယား) နဲ့ (၇) Bayesian Estimation (ဘေးွစ်ယန်း ခန့်မှန်းခြင်း) တို့ပဲ ဖြစ်ပါတယ်။

ဒီအခန်းရဲ့အဓိက ရည်ရွယ်ချက်က **ဒေတာကနေ ဘယ်လို ကောက်ချက်ချမလဲ** ဆိုတဲ့ ခန့်မှန်းခြေရဲ့ အခြေခံတွေကို နားလည်ဖို့ပါ။

၆.၁ ဖော်ပြချက်စာရင်းဇယား (Descriptive Statistics)

ဒီအပိုင်းမှာ ဒေတာတွေကို ဘယ်လို အနှစ်ချုပ် ဖော်ပြသလဲဆိုတာ လေ့လာရမှာပါ။

- **Frequency Table (ကြိမ်နှုန်းဇယား)**
  ဒေတာတွေကို အတန်းအစားတွေ (Classes) ခွဲပြီး တစ်တန်းစီမှာ ဘယ်နှစ်ခုပါသလဲဆိုတဲ့ ဇယားပါ။ ဒါက ဒေတာရဲ့ပုံသဏ္ဌာန်ကို မြင်သာအောင် လုပ်ပေးပါတယ်။

- **Histogram (ပြောင်းလဲမှုပုံ)**
  Frequency Table ကို ပုံဖော်တဲ့ နည်းလမ်းပါ။ အတန်းတစ်ခုစီကို ဘားတစ်ခုအနေနဲ့ ဆွဲပြီး အမြင့်က ကြိမ်နှုန်းကို ကိုယ်စားပြုပါတယ်။ Relative Frequency Histogram ဆိုရင် ဘားတစ်ခုစီရဲ့ဧရိယာက သက်ဆိုင်ရာ ရာခိုင်နှုန်းကို ကိုယ်စားပြုပါတယ်။

- **Sample Mean (နမူနာပျမ်းမျှ - $\bar{x}$)**
  ဒေတာအစုတစ်ခုရဲ့ပျမ်းမျှတန်ဖိုးပါ။

  - $\bar{x} = (\Sigma x_i) / n$

- **Sample Variance (နမူနာကွဲလွဲမှု - s^2)**

  ဒေတာတွေ ပျမ်းမျှကနေ ဘယ်လောက်ကွာဝေးသလဲဆိုတဲ့ အတိုင်းအတာပါ။

  - $s^2 = \Sigma (x_i - \bar{x})^2 / (n-1)$ (ဒီမှာ n-1 နဲ့ စားတာက Unbiased Estimator ဖြစ်ဖို့ပါ)

- **Sample Standard Deviation (နမူနာစံသွေဖည်မှု - s)**

  $s = \sqrt{s^2}$

၆.၂ စူးစမ်းရှာဖွေရေး ဒေတာခွဲခြမ်းစိတ်ဖြာခြင်း (Exploratory Data Analysis)

ဒီအပိုင်းမှာ ဒေတာရဲ့ ပုံသဏ္ဌာန်နဲ့ အထူးအချက်တွေကို ရှာဖွေတဲ့ နည်းလမ်းတွေကို လေ့လာရမှာပါ။

- **Stem-and-Leaf Display (ပင်စည်နဲ့အရွက် ဖော်ပြမှု)**

  ဒေတာတွေကို အစီအစဉ်တကျ ပြသပြီး မူရင်းတန်ဖိုးတွေကို မဆုံးရှုံးစေတဲ့ နည်းလမ်းပါ။ Histogram လိုပဲ ပုံသဏ္ဌာန်ကို မြင်ရပြီး တန်ဖိုးတွေကိုလည်း ဖတ်လို့ရပါတယ်။

- **Five-Number Summary (ဂဏန်းငါးလုံးအကျဉ်းချုပ်)**

  ဒေတာတစ်ခုရဲ့ အခြေခံအကျဆုံး သွင်ပြင်လက္ခဏာ (၅) ခုကို ဖော်ပြတာပါ။

  1. Minimum (အနည်းဆုံး)
  2. First Quartile (Q_1) - ၂၅ ရာခိုင်နှုန်းမြောက် တန်ဖိုး
  3. Median (Q_2) - ၅၀ ရာခိုင်နှုန်းမြောက် တန်ဖိုး
  4. Third Quartile (Q_3) - ၇၅ ရာခိုင်နှုန်းမြောက် တန်ဖိုး
  5. Maximum (အများဆုံး)

- **Box Plot (သေတ္တာပုံ)**

  Five-Number Summary ကို ပုံဖော်တဲ့ နည်းလမ်းပါ။ ဒါက ဒေတာရဲ့ ပျံ့နှံ့မှု၊ အချိုးကျမှု (Skewness) နဲ့ ထူးခြားတဲ့ တန်ဖိုးတွေ (Outliers) ကို မြင်သာအောင် ပြသပါတယ်။

- **Outliers (ထူးခြားသောတန်ဖိုး)**
  တခြားတန်ဖိုးတွေနဲ့ အများကြီး ကွာဝေးတဲ့ တန်ဖိုးတွေပါ။ ဒါတွေက မှတ်တမ်းမှားခြင်း ဒါမှမဟုတ် အရေးကြီးတဲ့ အချက်အလက်တွေ ဖြစ်နိုင်ပါတယ်။

၆.၃ အစီအစဉ်စာရင်းဇယား (Order Statistics)

ဒီအပိုင်းမှာ ဒေတာတွေကို အငယ်ဆုံးကနေ အကြီးဆုံးအထိ စီပြီး သူတို့ရဲ့ဖြန့်ဖြူးမှုကို လေ့လာရမှာပါ။

- **အဓိပ္ပါယ်**
  $X_1, X_2, \ldots, Xn$ ကို အငယ်ဆုံးကနေ အကြီးဆုံးအထိ စီထားတဲ့ အစီအစဉ်ကို Order Statistics လို့ ခေါ်ပြီး $Y_1 < Y_2 < \ldots < Yn$ လို့ ရေးပါတယ်။

- **အရေးပါမှု**
  - နမူနာပျမ်းမျှ (Median) က Y_{(n+1)/2} ဖြစ်ပြီး n စုံဆိုရင် နှစ်ခုရဲ့ ပျမ်းမျှဖြစ်ပါတယ်။
  - နမူနာနယ်ပယ် (Range) က $Yn - Y_1$ ဖြစ်ပါတယ်။

- **Distribution of Order Statistics**
  r-th Order Statistic Yr ရဲ့ pdf က အောက်ပါအတိုင်းဖြစ်ပါတယ်။
  - g_r(y) = n! / [(r-1)!(n-r)!] * [F(y)]^{r-1} * [1-F(y)]^{n-r} * f(y)

၆.၄ အများဆုံးဖြစ်နိုင်ခြေခန့်မှန်းခြင်း (Maximum Likelihood Estimation - MLE)

ဒါက Parameter တွေကို ခန့်မှန်းတဲ့ နည်းလမ်းတွေထဲက အား အကြီးဆုံးနဲ့ အသုံးအများဆုံး နည်းလမ်းပါ။

- **Likelihood Function (ဖြစ်နိုင်ခြေဖွင့်ဆိုချက်)**
  ဒါက ကျွန်တော်တို့ မြင်ရတဲ့ ဒေတာတွေရဲ့ဖြစ်တန်စွမ်းကို Parameter θ ရဲ့ လုပ်ဆောင်ချက်အနေနဲ့ ပြန်ရေးတာပါ။

- $L(\theta) = f(x_1;\theta) * f(x_2;\theta) * \dots * f(xn;\theta)$

- **MLE ကို ရှာတဲ့ နည်းလမ်း**

  L(θ) ကို အကြီးဆုံးဖြစ်စေတဲ့ θ တန်ဖိုးက MLE ဖြစ်ပါတယ်။ များသောအားဖြင့် ln L(θ) ကို အကြီးဆုံးဖြစ်အောင် ရှာတာက ပိုလွယ်ပါတယ်။ (ဘာကြောင့်လဲဆိုတော့ log က monotonic function ဖြစ်လို့)

- **ဥပမာ - Normal Distribution ရဲ့ MLE**

  $X_1, X_2, \dots, Xn$ က $N(\mu, \sigma^2)$ ကနေ Random Sample ဆိုရင် MLE တွေက

  - $\mu \hat{} = \bar{X}$ (နမူနာပျမ်းမျှ)
  - $\sigma^{2} \hat{} = (1/n) \Sigma (X_i - \bar{X})^2$ (ဒါက Biased Estimator ဖြစ်ပြီး Unbiased ဖြစ်ဖို့ n-1 နဲ့စားတဲ့ S^2 ကို သုံးပါတယ်)

- **Unbiased Estimator (ဘက်မလိုက်သော ခန့်မှန်းချက်)**

  Estimator တစ်ခုရဲ့ မျှော်မှန်းချက် (Expected Value) က Parameter ရဲ့ တန်ဖိုးနဲ့ ညီရင် အဲဒီ Estimator က Unbiased ဖြစ်ပါတယ်။ $E(\hat{\theta}) = \theta$

၆.၅ ရိုးရှင်းသော ဆုတ်ယုတ်ခွဲခြမ်းစိတ်ဖြာခြင်း (A Simple Regression Problem)

ဒီအပိုင်းမှာ ကိန်းရှင်နှစ်ခုကြား ဆက်စပ်မှုကို မော်ဒယ်တစ်ခုနဲ့ ဘယ်လို ခန့်မှန်းသလဲဆိုတာ လေ့လာရမှာပါ။

- **Linear Regression Model (မျဉ်းဖြောင့်ဆုတ်ယုတ်မှုမော်ဒယ်)**

  Y က X ရဲ့ မျဉ်းဖြောင့်လုပ်ဆောင်ချက်တစ်ခုဖြစ်ပြီး Error ပါဝင်တယ်လို့ ယူဆပါတယ်။

  - $Y_i = \alpha + \beta x_i + \varepsilon_i$ ($\varepsilon_i \sim N(0, \sigma^2)$)

- **Least Squares Estimates (အနည်းဆုံးနှစ်ထပ်ကိန်းခန့်မှန်းချက်များ)**

  α နဲ့ β ကို ခန့်မှန်းတဲ့ နည်းလမ်းပါ။ Residuals ($Y_i - \hat{Y}_i$) တွေရဲ့ နှစ်ထပ်ကိန်းပေါင်းလဒ်ကို အနည်းဆုံးဖြစ်အောင် လုပ်တာပါ။

  - $\beta \hat{} = \Sigma (x_i - \bar{x})(Y_i - \bar{Y}) / \Sigma (x_i - \bar{x})^2$

- $\alpha \hat{=} \bar{Y} - \beta\bar{x}$

၆.၆ Maximum Likelihood Estimator တွေရဲ့ Asymptotic Distribution

ဒီအပိုင်းမှာ MLE တွေရဲ့ နမူနာအရွယ်အစား ကြီးလာတဲ့အခါ ဖြန့်ဖြူးမှုပုံစံကို လေ့လာရမှာပါ။

- **အရေးကြီးတဲ့ ရလဒ်**
  ပုံမှန် အခြေအနေတွေမှာ MLE θ ရဲ့ဖြန့်ဖြူးမှုက n ကြီးလာတာနဲ့အမျှ ပုံမှန်ဖြန့်ဖြူးမှုဆီကို ချဉ်းကပ်လာပါတယ်။
  - $\theta \approx N(\theta, 1 / [-n * E(\partial^2 \ln f / \partial\theta^2)])$
- **Rao-Cramér Lower Bound**
  ဒါက Unbiased Estimator တစ်ခုရဲ့ Variance အတွက် အနိမ့်ဆုံး ကန့်သတ်ချက်ပါ။ MLE ရဲ့ Variance က ဒီကန့်သတ်ချက်ဆီကို ချဉ်းကပ်လာပါတယ်။

၆.၇ လုံလောက်သော စာရင်းဇယား (Sufficient Statistics)

ဒီအပိုင်းမှာ ဒေတာတွေထဲက Parameter နဲ့ ဆိုင်တဲ့ အချက်အလက်အားလုံးကို သယ်ဆောင်တဲ့ စာရင်းဇယားတွေအကြောင်း လေ့လာရမှာပါ။

- **အဓိပ္ပါယ် (Factorization Theorem)**
  Y က θ အတွက် Sufficient Statistic ဖြစ်ဖို့ Joint pdf/pmf ကို အောက်ပါအတိုင်း ခွဲလို့ရရပါမယ်။
  - $f(x_1, x_2, \ldots, xn; \theta) = g(Y; \theta) * h(x_1, x_2, \ldots, xn)$
  - ဒီမှာ g က Y ကတစ်ဆင့် θ ပေါ်မူတည်ပြီး h က θ ပေါ် မူတည်မနေဘူး။
- **အရေးပါမှု**
  Sufficient Statistic က ဒေတာထဲက Parameter နဲ့ ဆိုင်တဲ့ အချက်အလက်အားလုံးကို ပါဝင်ပါတယ်။ ဒါကြောင့် အကောင်းဆုံး ခန့်မှန်းချက်တွေက Sufficient Statistic ရဲ့ လုပ်ဆောင်ချက်တွေ ဖြစ်ပါတယ်။

၆.၈ ဘေးွစ်ယန်း ခန့်မှန်းခြင်း (Bayesian Estimation)

ဒါက Parameter ကို ကျပန်းကိန်းရှင်တစ်ခုအနေနဲ့ သတ်မှတ်ပြီး ခန့်မှန်းတဲ့ နည်းလမ်းပါ။

- **Prior Distribution (ကြိုတင်ဖြန့်ဖြူးမှု)**
  Parameter θ ရဲ့ ဒေတာမကြည့်ခင် ဖြန့်ဖြူးမှုကို ကိုယ်စားပြုပါတယ်။ h(θ)
- **Posterior Distribution (နောက်ဆက်တွဲဖြန့်ဖြူးမှု)**
  ဒေတာကို မြင်ပြီးနောက် θ ရဲ့ဖြန့်ဖြူးမှုပါ။ Bayes' Theorem နဲ့ တွက်ပါတယ်။
  - k(θ | x) ∝ h(θ) * L(θ) (L(θ) က Likelihood Function)
- **Bayes Estimator**
  Posterior Distribution ရဲ့ Mean (သို့) Median ကို Bayes Estimator အနေနဲ့ သုံးပါတယ်။
  (Loss Function ပေါ်မူတည်ပြီး ကွာခြားနိုင်ပါတယ်)

အချုပ်အားဖြင့် ဒီအခန်းက ဒေတာတွေကနေ လူဦးရေရဲ့သွင်ပြင်လက္ခဏာတွေကို ခန့်မှန်းတဲ့ နည်းလမ်းတွေကို သင်ကြားပေးထားပါတယ်။ Descriptive Statistics နဲ့ ဒေတာကို နားလည်ခြင်းကနေ စတင်ပြီး MLE လိုမျိုး အားကောင်းတဲ့ ခန့်မှန်းနည်းလမ်းတွေ၊ Sufficient Statistics ရဲ့ အယူအဆ၊ Bayesian ခန့်မှန်းခြင်းအထိ အဆင့်ဆင့် တိုးတက်သွားပါတယ်။ ဒါတွေအားလုံးက နောက်အခန်းတွေမှာ သင်ယူရမယ့် Interval Estimation နဲ့ Hypothesis Testing ရဲ့ အခြေခံအုတ်မြစ် ဖြစ်ပါတယ်။

အခန်း (၇) ကြားကာလခန့်မှန်းခြင်း (Interval Estimation)

ဒီအခန်းရဲ့အဓိက ရည်ရွယ်ချက်က **လူဦးရေရဲ့သွင်ပြင်လက္ခဏာ (Parameter) တစ်ခုကို တစ်ခုတည်းသော တန်ဖိုးနဲ့ ခန့်မှန်းတာမျိုး (Point Estimation) မဟုတ်ဘဲ၊ တန်ဖိုးတစ်ခုရဲ့ အကွာအဝေး (Interval) နဲ့ ခန့်မှန်းတာကို Interval Estimation လို့ ခေါ်ပါတယ်။** ဒီအခန်းမှာ သင်ယူရမှာက (၁) Confidence Intervals for Means (ပျမ်းမျှအတွက် ယုံကြည်မှုကြားကာလ)၊ (၂) ပျမ်းမျှနှစ်ခု ကွာခြားချက်အတွက် ယုံကြည်မှုကြားကာလ၊ (၃) အချိုးအစား (Proportion) အတွက် ယုံကြည်မှုကြားကာလ၊ (၄) နမူနာအရွယ်အစား သတ်မှတ်ခြင်း၊ (၅) ရာခိုင်နှုန်းအတွက် ဖြန့်ဖြူးမှုမလိုအပ်သော ယုံကြည်မှုကြားကာလ၊ (၆) ဆုတ်ယုတ်ခွဲခြမ်းစိတ်ဖြာခြင်းဆိုင်ရာ အကြောင်းအရာများနဲ့ (၇) ပြန်လည်နမူနာယူခြင်းနည်းလမ်းများ (Resampling Methods) တို့ပဲ ဖြစ်ပါတယ်။

ဒီအခန်းရဲ့အဓိက ရည်ရွယ်ချက်က **ကျွန်တော်တို့ရဲ့ခန့်မှန်းချက်တွေ ဘယ်လောက် ယုံကြည်ရလဲဆိုတာကို တိုင်းတာဖို့ပါ။**

၇.၁ ပျမ်းမျှအတွက် ယုံကြည်မှုကြားကာလ (Confidence Intervals for Means)

ဒီအပိုင်းမှာ လူဦးရေရဲ့ပျမ်းမျှ (μ) အတွက် ယုံကြည်မှုကြားကာလကို ဘယ်လို တွက်သလဲဆိုတာ လေ့လာရမှာပါ။

- **အခြေခံ အယူအဆ**
  နမူနာပျမ်းမျှ $\bar{X}$ ရဲ့ဖြန့်ဖြူးမှုကို သုံးပြီး μ ပါဝင်နိုင်တဲ့ အကွာအဝေးကို ဆောက်တာပါ။
- **σ သိရင် (Variance သိရင်)**
  $\bar{X}$ က $N(\mu, \sigma^2/n)$ ဖြစ်တဲ့အတွက် $Z = (\bar{X} - \mu) / (\sigma/\sqrt{n})$ က $N(0,1)$ ဖြစ်ပါတယ်။ ဒါကြောင့် μ အတွက် (1-α)100% ယုံကြည်မှုကြားကာလက
  - $\bar{x} \pm z_{\alpha/2} * (\sigma/\sqrt{n})$
- **σ မသိရင် (Variance မသိရင်)**
  σ အစား Sample Standard Deviation s ကို သုံးပြီး $t = (\bar{X} - \mu) / (s/\sqrt{n})$ က t(n-

1) ဖြန့်ဖြူးမှုအတိုင်း ဖြစ်ပါတယ်။ ဒါကြောင့် μ အတွက် (1-α)100% ယုံကြည်မှုကြားကာလက

- $\bar{x} \pm t_{\{\alpha/2\}(n-1)} * (s/\sqrt{n})$

- **ဥပမာ - မီးသီးတစ်လုံးရဲ့ ပျမ်းမျှသက်တမ်း**
  မီးသီး ၂၇ လုံးကို စမ်းသပ်တယ်။ ပျမ်းမျှသက်တမ်း $\bar{x}$ = 1478 နာရီ၊ စံသွေဖည်မှု σ = 36 နာရီ (သိပြီးသား) ဆိုရင် ၉၅% ယုံကြည်မှုကြားကာလက 1478 ± 1.96*(36/√27) ဖြစ်ပြီး [1464.42, 1491.58] ရပါတယ်။

၇.၂ ပျမ်းမျှနှစ်ခု ကွာခြားချက်အတွက် ယုံကြည်မှုကြားကာလ

ဒီအပိုင်းမှာ လူဦးရေနှစ်ခုရဲ့ ပျမ်းမျှ ကွာခြားချက် ($\mu_1 - \mu_2$) အတွက် ယုံကြည်မှုကြားကာလကို ဘယ်လို တွက်သလဲဆိုတာ လေ့လာရမှာပါ။

- **အခြေအနေ (၃) မျိုး**

  1. **Variance နှစ်ခုလုံး သိရင်:** $Z = (\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2) / \sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}$ ကို သုံးပါတယ်။
  2. **Variance နှစ်ခုလုံး မသိပေမယ့် ညီမျှတယ်လို့ ယူဆရင်:** Pooled Variance sp^2 ကို သုံးပြီး t ဖြန့်ဖြူးမှုကို သုံးပါတယ်။
  3. **Variance နှစ်ခုလုံး မသိဘဲ မညီမျှနိုင်ဘူးလို့ ယူဆရင်:** Welch's t-test ကို သုံးပြီး အနီးစပ်ဆုံး ယုံကြည်မှုကြားကာလကို တွက်ပါတယ်။

- **ဥပမာ - ကျောင်းသားနှစ်မျိုးရဲ့ စာမေးပွဲရမှတ်**
  ကျောင်းကြီးနဲ့ ကျောင်းငယ်က ကျောင်းသားတွေရဲ့ သင်္ချာအမှတ်တွေကို နှိုင်းယှဉ်တယ်။ ယုံကြည်မှုကြားကာလမှာ ၀ ပါရင် ပျမ်းမျှနှစ်ခု သိသိသာသာ မကွာဘူးလို့ ဆိုနိုင်ပါတယ်။

၇.၃ အချိုးအစားအတွက် ယုံကြည်မှုကြားကာလ (Confidence Intervals for Proportions)

ဒီအပိုင်းမှာ လူဦးရေရဲ့အချိုးအစား (p) အတွက် ယုံကြည်မှုကြားကာလကို ဘယ်လို တွက်သလဲဆိုတာ လေ့လာရမှာပါ။

- **အနီးစပ်ဆုံး ယုံကြည်မှုကြားကာလ (n ကြီးရင်)**
  p̂ = y/n (နမူနာအချိုးအစား) ရဲ့ဖြန့်ဖြူးမှုက ပုံမှန်နီးပါးဖြစ်တဲ့အတွက် (1-α)100% ယုံကြည်မှုကြားကာလက
  - $\hat{p} \pm z_{\alpha/2} * \sqrt{\hat{p}(1-\hat{p})/n}$
- **ဥပမာ - မဲဆန္ဒရှင်တစ်ဦးရဲ့ထောက်ခံမှု**
  မဲဆန္ဒရှင် ၃၅၁ ဦးကို စစ်တမ်းကောက်တယ်။ ၁၈၅ ဦးက ကိုယ်စားလှယ်လောင်းကို ထောက်ခံတယ်ဆိုရင် ၉၅% ယုံကြည်မှုကြားကာလက $0.527 \pm 1.96 * \sqrt{0.527*0.473/351}$ ဖြစ်ပြီး [0.475, 0.579] ရပါတယ်။

၇.၄ နမူနာအရွယ်အစား သတ်မှတ်ခြင်း (Sample Size)

ဒီအပိုင်းမှာ သတ်မှတ်ထားတဲ့ ယုံကြည်မှုအဆင့်နဲ့ တိကျမှုအတွက် နမူနာဘယ်လောက်ယူရမလဲဆိုတာ တွက်နည်းကို လေ့လာရမှာပါ။

- **Mean အတွက်**
  ယုံကြည်မှုကြားကာလရဲ့အမှားအယွင်း (Error) က $E = z_{\alpha/2} * (\sigma/\sqrt{n})$ ဖြစ်တဲ့အတွက် n ကို အောက်ပါအတိုင်း တွက်ပါတယ်။
  - $n = (z_{\alpha/2} * \sigma / E)^2$
- **Proportion အတွက်**
  $E = z_{\alpha/2} * \sqrt{p(1-p)/n}$ ဖြစ်တဲ့အတွက် p ကို မသိရင် အဆိုးဆုံး အခြေအနေ p = 0.5 ကို သုံးပြီး တွက်ပါတယ်။
  - $n = (z_{\alpha/2}^2 * p(1-p)) / E^2$

- **ဥပမာ - ရွေးကောက်ပွဲ စစ်တမ်း**
  မဲဆန္ဒရှင်တွေရဲ့ ထောက်ခံမှုကို ၉၅% ယုံကြည်မှုနဲ့ အမှား ၃% အတွင်း ခန့်မှန်းချင်ရင် n = $(1.96^2 * 0.5 * 0.5) / 0.03^2 \approx 1067$ ယူရပါမယ်။

၇.၅ ရာခိုင်နှုန်းအတွက် ဖြန့်ဖြူးမှုမလိုအပ်သော ယုံကြည်မှုကြားကာလ (Distribution-Free Confidence Intervals for Percentiles)

ဒီအပိုင်းမှာ ဖြန့်ဖြူးမှုပုံစံကို မသိရင်တောင် လူဦးရေရဲ့ ရာခိုင်နှုန်း (Percentile) အတွက် ယုံကြည်မှုကြားကာလကို ဘယ်လို ဆောက်သလဲဆိုတာ လေ့လာရမှာပါ။

- **အခြေခံ အယူအဆ**
  Order Statistics ကို သုံးပြီး ယုံကြည်မှုကြားကာလကို ဆောက်ပါတယ်။ ဥပမာ - လူဦးရေရဲ့ ပျမ်းမျှ (Median) m အတွက် $Y_i < m < Y_j$ ဖြစ်တဲ့ ဖြစ်တန်စွမ်းကို Binomial Distribution နဲ့ တွက်ပါတယ်။

- **ဥပမာ - ငါးတစ်ကောင်ရဲ့ အလယ်အလတ်အရှည်**
  ငါး ၉ ကောင်ရဲ့ အရှည်တွေကို စီထားတယ်။ ဒုတိယအငယ်ဆုံး Y_2 နဲ့ ဒုတိယအကြီးဆုံး Y_8 ကြားမှာ လူဦးရေရဲ့ ပျမ်းမျှအရှည် ပါဖို့ ဖြစ်တန်စွမ်းက $P(Y_2 < m < Y_8) = 0.961$ ဖြစ်ပါတယ်။ ဒါက ၉၆.၁% ယုံကြည်မှုကြားကာလကို ပေးပါတယ်။

၇.၆ ဆုတ်ယုတ်ခွဲခြမ်းစိတ်ဖြာခြင်းဆိုင်ရာ အကြောင်းအရာများ (More Regression)

ဒီအပိုင်းမှာ ဆုတ်ယုတ်ခွဲခြမ်းစိတ်ဖြာခြင်း (Regression) မှာ သုံးတဲ့ ယုံကြည်မှုကြားကာလတွေအကြောင်း လေ့လာရမှာပါ။

- **Slope β အတွက် ယုံကြည်မှုကြားကာလ**
  β ရဲ့ ဖြန့်ဖြူးမှုကို သုံးပြီး β အတွက် ယုံကြည်မှုကြားကာလကို တွက်ပါတယ်။

- **Prediction Interval (ကြိုတင်ခန့်မှန်းမှုကြားကာလ)**
  $X = x_0$ အတွက် Y ရဲ့ တန်ဖိုးကို ကြိုတင်ခန့်မှန်းတဲ့ ကြားကာလပါ။ ဒါက Confidence

Interval for Mean ထက် ပိုကျယ်ပါတယ် (ဘာကြောင့်လဲဆိုတော့ တစ်ခုချင်းစီရဲ့ကွဲလွဲမှုကို ထည့်တွက်လို့)။

၇.၇ ပြန်လည်နမူနာယူခြင်းနည်းလမ်းများ (Resampling Methods)

ဒီအပိုင်းမှာ ခေတ်မီတဲ့ ကွန်ပျူတာနည်းပညာကို သုံးပြီး ဖြန့်ဖြူးမှုကို ခန့်မှန်းတဲ့ နည်းလမ်းတွေအကြောင်း လေ့လာရမှာပါ။

- **Bootstrapping (ဘွတ်စထရပ်ပင်)**
  ရှိပြီးသား နမူနာဒေတာကနေ ထပ်ခါထပ်ခါ နမူနာယူပြီး (Re-sampling with replacement) စာရင်းဇယားတစ်ခုရဲ့ဖြန့်ဖြူးမှုကို ခန့်မှန်းတဲ့ နည်းလမ်းပါ။

- **ဘာကြောင့် အသုံးဝင်တာလဲ**
  သီအိုရီအရ ဖြန့်ဖြူးမှုကို ရှာရခက်တဲ့ စာရင်းဇယားတွေအတွက် အထူးအသုံးဝင်ပါတယ်။ ဥပမာ - Trimmed Mean လိုမျိုး စာရင်းဇယားတွေရဲ့ဖြန့်ဖြူးမှုကို ခန့်မှန်းဖို့ သုံးပါတယ်။

အချုပ်အားဖြင့် ဒီအခန်းက ခန့်မှန်းချက်တွေရဲ့တိကျမှုကို တိုင်းတာတဲ့ နည်းလမ်းတွေကို သင်ကြားပေးထားပါတယ်။ Confidence Intervals တွေက ကျွန်တော်တို့ရဲ့ခန့်မှန်းချက်တွေ ဘယ်လောက် ယုံကြည်ရလဲဆိုတာ ပြောပြပြီး၊ Sample Size တွက်နည်းက လေ့လာမှုတစ်ခုကို ဘယ်လောက်ဒေတာယူရမလဲဆိုတာ ဆုံးဖြတ်ပေးပါတယ်။ Distribution-Free နည်းလမ်းတွေနဲ့ Bootstrapping က သီအိုရီအရ ခက်ခဲတဲ့ အခြေအနေတွေအတွက် လက်တွေ့ကျတဲ့ အဖြေတွေ ပေးပါတယ်။ ဒါတွေအားလုံးက နောက်အခန်းတွေမှာ သင်ယူရမယ့် Hypothesis Testing ရဲ့ အခြေခံအုတ်မြစ် ဖြစ်ပါတယ်။

အခန်း (၈) စာရင်းအင်းဆိုင်ရာ ယူဆချက်များကို စစ်ဆေးခြင်း

ဒီအခန်းရဲ့အဓိက ရည်ရွယ်ချက်က **လူဦးရေတစ်ခုရဲ့သွင်ပြင်လက္ခဏာ (Parameter) နဲ့ ပတ်သက်ပြီး ယူဆထားတဲ့ အဆိုပြုချက် (Hypothesis) တစ်ခုကို ဒေတာတွေသုံးပြီး ဘယ်လို စစ်ဆေးမလဲ (Tests of Statistical Hypotheses)** ဆိုတာကို လေ့လာဖို့ပါ။

ဒီအခန်းမှာ သင်ယူရမှာက (၁) ပျမ်းမျှတစ်ခုအကြောင်း စစ်ဆေးခြင်း (Tests About One Mean)၊ (၂) ပျမ်းမျှနှစ်ခု ညီမျှမှုကို စစ်ဆေးခြင်း (Tests of the Equality of Two Means)၊ (၃) အချိုးအစားများအကြောင်း စစ်ဆေးခြင်း (Tests About Proportions)၊ (၄) Wilcoxon Tests (ဖြန့်ဖြူးမှုပုံစံမလိုအပ်သော စစ်ဆေးမှုများ)၊ (၅) စစ်ဆေးမှုတစ်ခုရဲ့ပါဝါ (Power of a Statistical Test)၊ (၆) အကောင်းဆုံး ငြင်းဆိုသည့်နယ်ပယ် (Best Critical Regions) နဲ့ (၇) Likelihood Ratio Tests တို့ပဲ ဖြစ်ပါတယ်။

ဒီအခန်းရဲ့အဓိက ရည်ရွယ်ချက်က **ဒေတာတွေကို သုံးပြီး ဆုံးဖြတ်ချက်တွေ ဘယ်လို ချမလဲ** ဆိုတဲ့ ဆုံးဖြတ်ချက်ချခြင်းရဲ့အခြေခံတွေကို နားလည်ဖို့ပါ။

၈.၁ ပျမ်းမျှတစ်ခုအကြောင်း စစ်ဆေးခြင်း (Tests About One Mean)

ဒီအပိုင်းမှာ လူဦးရေတစ်ခုရဲ့ပျမ်းမျှ (μ) နဲ့ ပတ်သက်တဲ့ ယူဆချက်တွေကို ဘယ်လို စစ်ဆေးသလဲဆိုတာ လေ့လာရမှာပါ။

- **အခြေခံ အယူအဆများ**
    - **Null Hypothesis (H_0) - ငြင်းဆိုလိုသော အဆိုပြုချက်:** စစ်ဆေးလိုတဲ့ ယူဆချက် (ဥပမာ - ပျမ်းမျှက ၅၀ ဖြစ်တယ် $H_0: \mu = 50$)။
    - **Alternative Hypothesis (H_1) - အစားထိုးအဆိုပြုချက်:** H_0 ကို ငြင်းဆိုလိုက်ရင် လက်ခံမယ့် ယူဆချက် (ဥပမာ - ပျမ်းမျှက ၅၀ ထက် ကြီးတယ် $H_1: \mu > 50$)။
    - **Test Statistic (စစ်ဆေးမှုစာရင်းဇယား):** ဒေတာတွေကနေ တွက်တဲ့ စာရင်းဇယားတစ်ခုပါ။ (ဥပမာ - $z = (\bar{x} - \mu_0) / (\sigma/\sqrt{n})$)

- **Critical Region (CR) - ငြင်းဆိုသည့်နယ်ပယ်:** Test Statistic ရဲ့တန်ဖိုးတွေထဲက H_0 ကို ငြင်းဆိုမယ့် အကွာအဝေးပါ။
- **p-value:** H_0 မှန်တယ်ဆိုရင် လက်ရှိရလဒ်ထက် ပိုပြီး စွန်းထွက်တဲ့ ရလဒ်မျိုး ရဖို့ ဖြစ်တန်စွမ်းပါ။ p-value ငယ်လေလေ H_0 ကို ငြင်းဆိုဖို့ ပိုခိုင်လုံလေပါ။

- **အမှားအယွင်း အမျိုးအစားများ**
  - **Type I Error (အမျိုးအစား ၁ အမှား):** H_0 မှန်ပေမယ့် ငြင်းဆိုမိတဲ့ အမှားပါ။ ဒီအမှားရဲ့ဖြစ်တန်စွမ်းကို α (Significance Level) လို့ ခေါ်ပါတယ်။
  - **Type II Error (အမျိုးအစား ၂ အမှား):** H_0 မှားပေမယ့် ငြင်းဆိုဖို့ ပျက်ကွက်တဲ့ အမှားပါ။ ဒီအမှားရဲ့ဖြစ်တန်စွမ်းကို β လို့ ခေါ်ပါတယ်။

- **စစ်ဆေးမှု အမျိုးအစားများ**
  - **σ သိရင် (Z-test):** $z = (\bar{x} - \mu_0) / (\sigma/\sqrt{n})$ ကို သုံးပါတယ်။
  - **σ မသိရင် (t-test):** $t = (\bar{x} - \mu_0) / (s/\sqrt{n})$ ကို သုံးပြီး $t(n-1)$ ဖြန့်ဖြူးမှုနဲ့ နှိုင်းယှဉ်ပါတယ်။

- **ဥပမာ - ကျောင်းသားတွေရဲ့ပျမ်းမျှ IQ**
  ကျောင်းသားတွေရဲ့ပျမ်းမျှ IQ က ၁၁၀ လား $H_0: \mu = 110$ ဆိုတာ စစ်ချင်တယ်။
  ကျောင်းသား ၁၆ ယောက်ရဲ့ပျမ်းမျှ $\bar{x} = 113.5$ ရတယ်။ $\sigma = 10$ လို့ သိရင် $z = (113.5 - 110) / (10/4) = 1.4$ ဖြစ်ပြီး $z_{0.05} = 1.645$ ထက် ငယ်တဲ့အတွက် H_0 ကို ငြင်းဆိုလို့ မရပါဘူး။

၈.၂ ပျမ်းမျှနှစ်ခု ညီမျှမှုကို စစ်ဆေးခြင်း (Tests of the Equality of Two Means)

ဒီအပိုင်းမှာ လူဦးရေနှစ်ခုရဲ့ပျမ်းမျှတွေ ညီမျှမှုကို ဘယ်လို စစ်ဆေးသလဲဆိုတာ လေ့လာရမှာပါ။

- **Two-Sample t-test (Variance နှစ်ခုလုံး မသိဘူး၊ ဒါပေမယ့် ညီမျှတယ်လို့ ယူဆရင်)**
  - $t = (\bar{x}_1 - \bar{x}_2) / (sp * \sqrt{(1/n_1 + 1/n_2)})$
  - ဒီမှာ $sp^2 = [(n_1-1)s_1^2 + (n_2-1)s_2^2] / (n_1+n_2-2)$ က Pooled Variance ပါ။

- ဒီ t က $t(n_1+n_2-2)$ ဖြန့်ဖြူးမှုအတိုင်း ဖြစ်ပါတယ်။

- **ဥပမာ - ဆေးနှစ်မျိုးရဲ့ထိရောက်မှု**
  ဆေးအသစ်နဲ့ ဆေးဟောင်းရဲ့ပျမ်းမျှထိရောက်မှု ကွာခြားသလား စစ်ချင်တယ်။ t-test ကို သုံးပြီး p-value ကို တွက်ပါတယ်။ p-value က α ထက် ငယ်ရင် ဆေးနှစ်မျိုးရဲ့ထိရောက်မှု ကွာခြားတယ်လို့ ဆုံးဖြတ်ပါတယ်။

၈.၃ အချိုးအစားများအကြောင်း စစ်ဆေးခြင်း (Tests About Proportions)

ဒီအပိုင်းမှာ လူဦးရေရဲ့အချိုးအစား (p) နဲ့ ပတ်သက်တဲ့ ယူဆချက်တွေကို ဘယ်လို စစ်ဆေးသလဲဆိုတာ လေ့လာရမှာပါ။

- **One-Sample Test for Proportion**
  - $z = (\hat{p} - p_0) / \sqrt{(p_0(1-p_0)/n)}$ ကို သုံးပြီး N(0,1) နဲ့ နှိုင်းယှဉ်ပါတယ်။

- **Two-Sample Test for Proportions**
  - $z = (\hat{p}_1 - \hat{p}_2) / \sqrt{(\hat{p}(1-\hat{p})(1/n_1 + 1/n_2))}$ ကို သုံးပါတယ်။ (ဒီမှာ $\hat{p} = (y_1+y_2)/(n_1+n_2)$ က Pooled Proportion ပါ)

- **ဥပမာ - မဲဆန္ဒရှင်တစ်ဦးရဲ့ထောက်ခံမှု**
  ကိုယ်စားလှယ်လောင်းတစ်ဦးရဲ့ထောက်ခံမှု ၅၀% ထက် ပိုလား H_0: p = 0.5 ဆိုတာ စစ်ချင်တယ်။ မဲဆန္ဒရှင် ၄၀၀ မှာ ၂၂၀ က ထောက်ခံတယ်ဆိုရင် z = (0.55 - 0.5) / √(0.5*0.5/400) = 2 ဖြစ်ပြီး z_{0.05} = 1.645 ထက် ကြီးတဲ့အတွက် H_0 ကို ငြင်းဆိုလို့ရပါတယ်။

၈.၄ Wilcoxon Tests (ဖြန့်ဖြူးမှုပုံစံမလိုအပ်သော စစ်ဆေးမှုများ)

ဒီအပိုင်းမှာ ဒေတာတွေ ပုံမှန်ဖြန့်ဖြူးမှု (Normal Distribution) အတိုင်း မဟုတ်ရင်တောင် သုံးလို့ရတဲ့ စစ်ဆေးမှုနည်းလမ်းတွေအကြောင်း လေ့လာရမှာပါ။

- **ဘာကြောင့် သုံးတာလဲ**
  ဒေတာတွေက ပုံမှန်မဟုတ်ရင် (ဥပမာ - အလွန်အမင်း စောင်းနေရင်) t-test ရဲ့ ရလဒ်တွေ ယုံကြည်ရမှု နည်းသွားနိုင်ပါတယ်။ Wilcoxon Test က ဒီလိုအခြေအနေမျိုးမှာ ပိုလုံခြုံပါတယ်။

- **Wilcoxon Signed-Rank Test (တစ်ခုတည်းသော နမူနာအတွက်)**
  ဒေတာတွေရဲ့ အမှတ်အသား (Sign) နဲ့ အဆင့် (Rank) ကို ပေါင်းပြီး စစ်ဆေးတဲ့ နည်းလမ်းပါ။

- **Wilcoxon Rank-Sum Test (နမူနာနှစ်ခုအတွက်)**
  နမူနာနှစ်ခုကို ပေါင်းပြီး အဆင့်သတ်မှတ်ကာ တစ်ခုရဲ့ အဆင့်ပေါင်းလဒ်ကို သုံးပြီး စစ်ဆေးတဲ့ နည်းလမ်းပါ။

၈.၅ စစ်ဆေးမှုတစ်ခုရဲ့ ပါဝါ (Power of a Statistical Test)

ဒီအပိုင်းမှာ စစ်ဆေးမှုတစ်ခုရဲ့ စွမ်းဆောင်ရည် (Power) ကို ဘယ်လို တိုင်းတာသလဲဆိုတာ လေ့လာရမှာပါ။

- **Power Function (ပါဝါဖွင့်ဆိုချက်)**
  H_0 မှားတဲ့အခါ H_0 ကို ငြင်းဆိုနိုင်တဲ့ ဖြစ်တန်စွမ်းကို Power = 1 - β လို့ သတ်မှတ်ပါတယ်။

- **အရေးပါမှု**
  Power များလေလေ စစ်ဆေးမှုက ပိုကောင်းလေလေပါ။ Power ကို မြှင့်တင်ဖို့ နမူနာအရွယ်အစား (n) ကို မြှင့်တင်နိုင်ပါတယ်။

၈.၆ အကောင်းဆုံး ငြင်းဆိုသည့်နယ်ပယ် (Best Critical Regions)

ဒီအပိုင်းမှာ သတ်မှတ်ထားတဲ့ Significance Level (α) အတွက် Power ကို အများဆုံးဖြစ်စေတဲ့ အကောင်းဆုံး စစ်ဆေးမှုကို ဘယ်လို ရှာသလဲဆိုတာ လေ့လာရမှာပါ။

- **Neyman-Pearson Lemma**
  ဒါက Simple Null Hypothesis နဲ့ Simple Alternative Hypothesis အတွက် အကောင်းဆုံး စစ်ဆေးမှုကို ရှာတဲ့ နည်းလမ်းပါ။ Likelihood Ratio $L(\theta_0)/L(\theta_1)$ ကို သုံးပြီး Critical Region ကို ဆောက်ပါတယ်။

၈.၇ Likelihood Ratio Tests

ဒီအပိုင်းမှာ Composite Hypothesis တွေအတွက် သုံးတဲ့ ယေဘုယျ စစ်ဆေးမှုနည်းလမ်းကို လေ့လာရမှာပါ။

- **အခြေခံ အယူအဆ**
  Likelihood Function ရဲ့ အမြင့်ဆုံးတန်ဖိုး (Maximum) ကို H_0 အောက်မှာရော H_1 အောက်မှာပါ တွက်ပြီး သူတို့ရဲ့ အချိုး (Ratio) ကို သုံးပါတယ်။
  - $\lambda = \max_{\theta \in \omega} L(\theta) / \max_{\theta \in \Omega} L(\theta)$ (ဒီမှာ ω က H_0 ရဲ့ Parameter Space၊ Ω က စုစုပေါင်း Parameter Space ပါ)
- **စစ်ဆေးမှု**
  λ က သိပ်ငယ်ရင် H_0 ကို ငြင်းဆိုပါတယ်။ ဒီနည်းလမ်းက t-test, F-test, Chi-square test စတဲ့ စစ်ဆေးမှုတွေရဲ့ အခြေခံ ဖြစ်ပါတယ်။

အချုပ်အားဖြင့် ဒီအခန်းက ဒေတာတွေကို သုံးပြီး ယူဆချက်တွေကို ဘယ်လို စစ်ဆေးမလဲဆိုတဲ့ အခြေခံတွေကို သင်ကြားပေးထားပါတယ်။ Hypothesis Testing ရဲ့ အခြေခံ အယူအဆတွေ (Null/Alternative Hypothesis, Type I/II Errors, p-value) ကနေ စတင်ပြီး ပျမ်းမျှ၊ အချိုးအစားတွေအတွက် စစ်ဆေးနည်းတွေ၊ ပုံမှန်မဟုတ်တဲ့ ဒေတာတွေအတွက် Wilcoxon Tests လိုမျိုး နည်းလမ်းတွေ၊ စစ်ဆေးမှုတစ်ခုရဲ့ စွမ်းဆောင်ရည်ကို တိုင်းတာတဲ့ Power နဲ့ အကောင်းဆုံး စစ်ဆေးမှုကို ရှာတဲ့ Neyman-Pearson Lemma အထိ အဆင့်ဆင့် တိုးတက်သွားပါတယ်။ ဒါတွေအားလုံးက နောက်အခန်းတွေမှာ သင်ယူရမယ့် ပိုမိုရှုပ်ထွေးတဲ့ စစ်ဆေးမှုတွေရဲ့ အခြေခံအုတ်မြစ် ဖြစ်ပါတယ်။

အခန်း (၉) နောက်ထပ် စစ်ဆေးမှုများ

ဒီအခန်းရဲ့ အဓိက ရည်ရွယ်ချက်က **ပိုမိုရှုပ်ထွေးတဲ့ အခြေအနေတွေအတွက် စာရင်းအင်းဆိုင်ရာ စစ်ဆေးမှုတွေ (More Tests)** ကို လေ့လာဖို့ပါ။ ဒီအခန်းမှာ သင်ယူရမှာက (၁) Chi-Square Goodness-of-Fit Tests (မော်ဒယ်ကောင်းမကောင်း စစ်ဆေးခြင်း)၊ (၂) Contingency Tables (အမျိုးအစားနှစ်ခုကြား ဆက်စပ်မှုစစ်ဆေးခြင်း)၊ (၃) One-Factor Analysis of Variance (ANOVA - တစ်ချက်ခွဲခြားမှုကွဲလွဲခွဲခြမ်းစိတ်ဖြာခြင်း)၊ (၄) Two-Way Analysis of Variance (နှစ်ချက်ခွဲခြားမှု ANOVA)၊ (၅) General Factorial နဲ့ 2^k Factorial Designs (အချက်ပေါင်းများစွာ စမ်းသပ်မှုပုံစံများ)၊ (၆) ဆုတ်ယုတ်ခွဲခြမ်းစိတ်ဖြာခြင်းနဲ့ ဆက်စပ်မှုဆိုင်ရာ စစ်ဆေးမှုများ (Tests Concerning Regression and Correlation) နဲ့ (၇) Statistical Quality Control (စာရင်းအင်းဆိုင်ရာ အရည်အသွေးထိန်းချုပ်ခြင်း) တို့ပဲ ဖြစ်ပါတယ်။

ဒီအခန်းရဲ့ အဓိက ရည်ရွယ်ချက်က **လက်တွေ့ဘဝမှာ ကြုံတွေ့ရတဲ့ ရှုပ်ထွေးတဲ့ ဒေတာတွေအတွက် စစ်ဆေးမှုနည်းလမ်းတွေကို နားလည်ဖို့ပါ။**

၉.၁ Chi-Square Goodness-of-Fit Tests (မော်ဒယ်ကောင်းမကောင်း စစ်ဆေးခြင်း)

ဒီအပိုင်းမှာ ကျွန်တော်တို့ရဲ့ ဒေတာတွေက ယူဆထားတဲ့ ဖြန့်ဖြူးမှု (Theoretical Distribution) နဲ့ ကိုက်ညီမှုရှိမရှိကို ဘယ်လို စစ်ဆေးသလဲဆိုတာ လေ့လာရမှာပါ။

- **အခြေခံ အယူအဆ**
  ဒေတာတွေကို အတန်းအစားတွေ (Categories) ခွဲပြီး လေ့လာတွေ့ရှိရတဲ့ ကြိမ်နှုန်း (Observed Frequency) နဲ့ ယူဆထားတဲ့ မော်ဒယ်အရ ရှိသင့်တဲ့ ကြိမ်နှုန်း (Expected Frequency) တို့ကို နှိုင်းယှဉ်တာပါ။

- **စစ်ဆေးမှုစာရင်းဇယား**
  $\chi^2 = \Sigma [(\text{Observed} - \text{Expected})^2 / \text{Expected}]$ ဖြစ်ပြီး ဒါက χ^2 ဖြန့်ဖြူးမှုအတိုင်း ဖြစ်ပါတယ်။

- Degrees of Freedom (df) = (အတန်းအစားအရေအတွက် - ၁ - ခန့်မှန်းထားတဲ့ ပါရာမီတာအရေအတွက်)

- **ဥပမာ - အန်စာတုံးတစ်ခု မျှတမှုရှိမရှိ စစ်ဆေးခြင်း**
  အန်စာတုံးတစ်ခုကို ၆၀ ကြိမ် လှိမ့်တယ်။ တစ်ဖက်စီကျတဲ့ အကြိမ်အရေအတွက်ကို ရေတွက်ပြီး မျှတရင် ဖြစ်သင့်တဲ့ အကြိမ်အရေအတွက် (၁၀ ကြိမ်) နဲ့ နှိုင်းယှဉ်ပါတယ်။ Chi-Square တန်ဖိုးက ကြီးရင် အန်စာတုံး မမျှတဘူးလို့ ဆုံးဖြတ်ပါတယ်။

၉.၂ Contingency Tables (အမျိုးအစားနှစ်ခုကြား ဆက်စပ်မှုစစ်ဆေးခြင်း)

ဒီအပိုင်းမှာ အမျိုးအစားနှစ်ခု (ဥပမာ - ကျား၊မ နဲ့ မဲပေးပုံ) ကြား ဆက်စပ်မှုရှိမရှိကို ဘယ်လို စစ်ဆေးသလဲဆိုတာ လေ့လာရမှာပါ။

- **အခြေခံ အယူအဆ**
  ဒေတာတွေကို ဇယားတစ်ခုအနေနဲ့ စီပြီး အမျိုးအစားတစ်ခုစီအတွက် လေ့လာတွေ့ရှိရတဲ့ ကြိမ်နှုန်းတွေကို မှတ်တမ်းတင်ပါတယ်။ ပြီးရင် အမျိုးအစားနှစ်ခု လွတ်လပ်တယ်ဆိုရင် ဖြစ်သင့်တဲ့ ကြိမ်နှုန်းတွေကို တွက်ပြီး Chi-Square စစ်ဆေးမှုကို လုပ်ပါတယ်။

- **ဥပမာ - ကျား၊မနဲ့ မဲပေးပုံ**
  ကျား၊မနဲ့ မဲပေးပုံ ဆက်စပ်မှုရှိလား စစ်ချင်တယ်။ Contingency Table တစ်ခု ဆောက်ပြီး Chi-Square စစ်ဆေးမှုကို လုပ်ပါတယ်။ Chi-Square တန်ဖိုးက ကြီးရင် ကျား၊မနဲ့ မဲပေးပုံ ဆက်စပ်မှုရှိတယ်လို့ ဆုံးဖြတ်ပါတယ်။

၉.၃ One-Factor Analysis of Variance (ANOVA - တစ်ချက်ခွဲခြားမှုကွဲလွဲခွဲခြမ်းစိတ်ဖြာခြင်း)

ဒီအပိုင်းမှာ အုပ်စုသုံးခု ဒါမှမဟုတ် သုံးခုထက်ပိုတဲ့ အုပ်စုတွေရဲ့ ပျမ်းမျှတွေ ညီမျှမှုကို ဘယ်လို စစ်ဆေးသလဲဆိုတာ လေ့လာရမှာပါ။ (t-test က အုပ်စုနှစ်ခုအတွက်ပဲ သုံးလို့ရတယ်)

- **အခြေခံ အယူအဆ**
  ဒေတာတွေရဲ့ စုစုပေါင်းကွဲလွဲမှု (Total Variation) ကို အပိုင်းနှစ်ပိုင်း ခွဲပါတယ်။

1. **Between-Group Variation (အုပ်စုကြားကွဲလွဲမှု):** အုပ်စုတစ်ခုနဲ့တစ်ခုကြား ပျမ်းမျှကွာခြားမှုကြောင့် ဖြစ်တဲ့ ကွဲလွဲမှု။
2. **Within-Group Variation (အုပ်စုအတွင်းကွဲလွဲမှု):** အုပ်စုတစ်ခုတည်းအတွင်းက ဒေတာတွေရဲ့ကွဲလွဲမှု။

- **F-test**
  - F = MSB / MSW (MSB = Between-Group Mean Square, MSW = Within-Group Mean Square)
  - F က ကြီးရင် အုပ်စုတွေရဲ့ပျမ်းမျှတွေ သိသိသာသာ ကွာခြားတယ်လို့ ဆုံးဖြတ်ပါတယ်။

- **ဥပမာ - ဆေးသုံးမျိုးရဲ့ထိရောက်မှု**
  ဆေးသုံးမျိုးရဲ့ပျမ်းမျှထိရောက်မှု ကွာခြားသလား စစ်ချင်တယ်။ ANOVA ကို သုံးပြီး F-test ကို လုပ်ပါတယ်။

၉.၄ Two-Way Analysis of Variance (နှစ်ချက်ခွဲခြားမှု ANOVA)

ဒီအပိုင်းမှာ အချက် (Factor) နှစ်ခုရဲ့သက်ရောက်မှုကို တစ်ပြိုင်နက် ဘယ်လို စစ်ဆေးသလဲဆိုတာ လေ့လာရမှာပါ။

- **အခြေခံ အယူအဆ**
  ဒေတာတွေရဲ့စုစုပေါင်းကွဲလွဲမှုကို အပိုင်း (၃) ပိုင်း ခွဲပါတယ်။
  1. **Factor A ကြောင့် ဖြစ်တဲ့ ကွဲလွဲမှု**
  2. **Factor B ကြောင့် ဖြစ်တဲ့ ကွဲလွဲမှု**
  3. **Interaction (အပြန်အလှန်သက်ရောက်မှု) ကြောင့် ဖြစ်တဲ့ ကွဲလွဲမှု** (Factor A ရဲ့ သက်ရောက်မှုက Factor B ရဲ့အဆင့်ပေါ်မူတည်ပြီး ပြောင်းလဲခြင်း)

- **F-tests**
  Factor A, Factor B နဲ့ Interaction အတွက် သီးခြား F-tests တွေ လုပ်ပါတယ်။

၉.၅ General Factorial နဲ့ 2^k Factorial Designs

ဒီအပိုင်းမှာ အချက် (Factor) သုံးခု ဒါမှမဟုတ် သုံးခုထက်ပိုတဲ့ စမ်းသပ်မှုတွေအတွက် ပုံစံတွေကို လေ့လာရမှာပါ။

- **Factorial Design**
  အချက်တိုင်းရဲ့အဆင့်တိုင်းကို အခြားအချက်တိုင်းရဲ့အဆင့်တိုင်းနဲ့ ပေါင်းစပ်ပြီး စမ်းသပ်တဲ့ နည်းလမ်းပါ။

- **2^k Factorial Design**
  အချက်တိုင်းကို အဆင့်နှစ်ဆင့်ပဲ (အနိမ့်နဲ့ အမြင့်) သတ်မှတ်ပြီး စမ်းသပ်တဲ့ နည်းလမ်းပါ။ ဒါက စမ်းသပ်မှုအရေအတွက်ကို သိသိသာသာ လျှော့ချပေးပါတယ်။

၉.၆ ဆုတ်ယုတ်ခွဲခြမ်းစိတ်ဖြာခြင်းနဲ့ ဆက်စပ်မှုဆိုင်ရာ စစ်ဆေးမှုများ

ဒီအပိုင်းမှာ ဆုတ်ယုတ်ခွဲခြမ်းစိတ်ဖြာခြင်း (Regression) နဲ့ ဆက်စပ်မှု (Correlation) အတွက် စစ်ဆေးမှုတွေကို လေ့လာရမှာပါ။

- **Slope (β) အတွက် စစ်ဆေးမှု**
  $t = \beta / SE(\beta)$ ကို သုံးပြီး $\beta = 0$ လား ဆိုတာ စစ်ပါတယ်။

- **Correlation Coefficient (ρ) အတွက် စစ်ဆေးမှု**
  $t = r * \sqrt{(n-2)} / \sqrt{(1-r^2)}$ ကို သုံးပြီး $\rho = 0$ လား ဆိုတာ စစ်ပါတယ်။

၉.၇ Statistical Quality Control (စာရင်းအင်းဆိုင်ရာ အရည်အသွေးထိန်းချုပ်ခြင်း)

ဒီအပိုင်းမှာ ထုတ်လုပ်မှုလုပ်ငန်းစဉ်တွေရဲ့အရည်အသွေးကို ထိန်းချုပ်ဖို့ စာရင်းအင်းနည်းလမ်းတွေကို ဘယ်လို သုံးသလဲဆိုတာ လေ့လာရမှာပါ။

- **Control Charts (ထိန်းချုပ်ဇယားများ)**
  လုပ်ငန်းစဉ်တစ်ခုရဲ့အခြေအနေကို အချိန်နဲ့အမျှ စောင့်ကြည့်တဲ့ ဂရပ်တွေပါ။

- **X̄ Chart:** နမူနာပျမ်းမျှတွေကို စောင့်ကြည့်ပြီး လုပ်ငန်းစဉ်ရဲ့ ပျမ်းမျှ ပြောင်းလဲမှုကို စစ်ဆေးတယ်။
- **R Chart:** နမူနာနယ်ပယ် (Range) တွေကို စောင့်ကြည့်ပြီး လုပ်ငန်းစဉ်ရဲ့ ကွဲလွဲမှု ပြောင်းလဲမှုကို စစ်ဆေးတယ်။
- **p Chart:** ချို့ယွင်းပစ္စည်းတွေရဲ့ အချိုးအစားကို စောင့်ကြည့်တယ်။
- **c Chart:** ပစ္စည်းတစ်ခုမှာရှိတဲ့ ချို့ယွင်းချက်အရေအတွက်ကို စောင့်ကြည့်တယ်။

- **Control Limits (ထိန်းချုပ်ကန့်သတ်ချက်များ)**
  - **Upper Control Limit (UCL):** အထက်ကန့်သတ်ချက်
  - **Center Line (CL):** ဗဟိုမျဉ်း
  - **Lower Control Limit (LCL):** အောက်ကန့်သတ်ချက်

- **အသုံးပြုပုံ**
  ဒေတာတစ်ခုက UCL ဒါမှမဟုတ် LCL ကို ကျော်သွားရင် လုပ်ငန်းစဉ်မှာ ပြဿနာရှိတယ်လို့ ဆိုလိုပြီး စစ်ဆေးမှုတွေ လုပ်ရပါမယ်။

အချုပ်အားဖြင့် ဒီအခန်းက လက်တွေ့ဘဝမှာ ကြုံတွေ့ရတဲ့ ရှုပ်ထွေးတဲ့ ဒေတာတွေအတွက် စစ်ဆေးမှုနည်းလမ်းတွေကို သင်ကြားပေးထားပါတယ်။ Chi-Square စစ်ဆေးမှုတွေကနေ စတင်ပြီး ANOVA လိုမျိုး အုပ်စုများစွာရဲ့ ပျမ်းမျှတွေကို နှိုင်းယှဉ်တဲ့ နည်းလမ်းတွေ၊ Factorial Designs လိုမျိုး အချက်ပေါင်းများစွာပါတဲ့ စမ်းသပ်မှုပုံစံတွေ၊ ဆုတ်ယုတ်ခွဲခြမ်းစိတ်ဖြာခြင်းနဲ့ ဆက်စပ်မှုအတွက် စစ်ဆေးမှုတွေ၊ နဲ့ ထုတ်လုပ်မှုလုပ်ငန်းစဉ်တွေရဲ့ အရည်အသွေးကို ထိန်းချုပ်ဖို့ Control Charts အထိ အကျုံးဝင်ပါတယ်။ ဒီအခန်းက စာရင်းအင်းပညာရပ်ရဲ့ လက်တွေ့အသုံးချပုံကို ကောင်းကောင်း သဘောပေါက်စေပါတယ်။